Advanced IPv6 Mobility Management for Next Generation Wireless Access Networks

Theodoros Pagtzis

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London.

Department of Computer Science

University College London

January 2005

'Εν αρχήν εί ο λόγος

Ιωάννης Βαπτιστής

Abstract

The theme of this thesis is the advancement of IP mobility management mechanisms to support robustly the delay requirements of interactive IP applications services. Current IPv6 mobility management standards are heavily dependent on reactive manipulations of the IP handoff process. Such modus operandi appears to be insufficient to address delay transparency during interactive IP communications. By means of proactive signalling deliberations with candidate points of attachment, the IP handoff management function can sustain delay transparency and, thus, enhance the perceived quality of communication during the mobile node's (MN's) next IP transition.

To this end, we assess to what extent recently proposed IPv6 macro-mobility management standards or alternative macro-/micro-mobility proposals meet such expectations. We extend our assessment by identifying experimentally factors that impede the performance of interactive applications over IPv6 Mobility standards, with particular focus on the process of IP handoffs. Such factors establish the control hypothesis of IP mobility management performance investigations.

Through experimental measurements we show that both the wireless network and terminal treat *reactive* IPv6 mobility as disconcerting disruption in interactive communications: IPv6 address configuration as a form of state establishment appears to impose unacceptable delays in packet flows communicated to the MN.

In this light, we investigate a novel IP mobility protocol architecture in support of delay seamlessness, irrespective of the MN's mobility pattern. Driven by experimental evidence, we reconsider the design of core parts of the IP mobility management function by: (i) giving special emphasis to delay transparency in the MN's communicated flows; (ii) scheduling core IP mobility component functions *proactively*, in advance of the MN's next IPv6 handoff. These functions are: (a) IP handoff management (b) IP flow forwarding. By means of efficient IP handoff management, the delay components in control of the IP layer can be eliminated. By means of efficient IP flow forwarding the mobility management function can circumvent delays incurred beyond the control of the network layer.

To effect such a form of mobility management, an efficient *routing neighbourhood discovery* algorithm is investigated. In this manner, the wireless network and the MN is aided with efficient forms of *mobility-hop* routing to aid the management of the MN's next IP handoff.

Neighbourhood discovery of mobility-hop routers is devised to facilitate proactive *state* establishment. Such a mechanism sets the basis for an open mobility management architecture in support of *any* ontological context of capabilities emanating from the network layer. Such an open design enables the evolution of IP handoff management from conventional signal-strength to abstract forms of context-aware, utility-based metrics, whereby diversity in selection of the MN's next point of IP attachment can be supported.

Ultimately, a comparative performance analysis of Proactive IPv6 mobility management versus the fast handoffs extension for MIPv6 (FMIPv6) is presented. This study identifies strengths as well as fundamental limitations for the two proposals.

By taking a structured approach in the design of the aforementioned proactive IPv6 mobility management architecture, this thesis advocates that *delay seamlessness can be addressed efficiently* in support of interactive IP application services. At the same time, it enhances the handoff management process by allowing advanced forms of *handoff diversity* to maximise the MN's service utility during its next IP transition.

Contents

List of Figures

List of Tables

Acknowledgements

A PhD study is a startling journey that transforms the person for ever. Irrespective of the amount of personal endeavour, however, such journey *cannot* come to fruition without the support of many people, that willingly (or unwillingly) influence its course and ultimately its outcome. To them all I am deeply grateful for being there.

First I would like to express my deepest appreciation to my parents Stylianos and Anastasia. It is due to their sacrifice, support and insight that I have managed to reach this far. With them I would like to extend my appreciation to the person that marked my life intellectually, Demosthenes Kotionis. I am most grateful to him for showing me the strength of analytical skills and above all for nurturing a feeling for excellence, while believing in me.

In the academic arena, I feel the urge to express my gratitude to my adviser, Prof. Peter T. Kirtsein for offering me the unique opportunity in pursuing and ultimately completing this research study next to him. Peter has been undoubtedly one of the Internet pioneers that *truly* carried his experience, insight and analytical thinking onto his research students as well as project efforts. Peter supported me through all my decisions with the utmost patience - virtue of a true adviser - and I am really grateful for that. It has been most certainly an honour and privilege to study under his guidance and direction. This research work is dedicated to him.

I would also like to thank my second adviser, Dr. Stephen Hailes, for his help and support. Steve always provided me with precious spartan feedback in this study, challenging the focus of my attention towards all the fine details that make research penetrating. I would like to express my gratitude to him for his patience with me and his persistence for excellence.

As an intern at Nokia Research Center (NRC), I had the pleasure of working with several outstanding researchers and people. I would like to thank Dr. Hannu Flinck for his guidance and invaluable discussions during this research effort. As my manager at NRC, Hannu gave me the opportunity to experience active research in one of the most lively research groups and familiarise myself with challenging research realities of a highly competitive environment. I would like to especially thank Charles E. Perkins for offering me a great opportunity in working under his guidance. Charles taught me with his silence, how to properly express my ideas and structure my research output. My collaboration with him has been an invaluable experience. I would also like to extend my appreciation to Rajeev Koodli. Rajeev showed me the importance of discipline and thoroughness in research as well as importance of implementation. On this front, I would also like to thank Cedric Westphal, Jari Malinen, T.J. Kniveton and Vijay Devarapalli for their invaluable discussions during this research as well as their friendship.

I would also like to thank Dr. Matthew and Marcia Allen, Dr. Hossam Affifi, Anders Nilsson and Tommi Yllamurto; they have been supportive, encouraging and above all cherished friends.

I am also truly indebted to Dr. Ken Carlberg, for his infinite kindness and brutal honesty in the review stages of this thesis. Ken influenced greatly the shape of this thesis through his critical, honest feedback on both research content as well as writing style and perspective in written research logos. I will treasure his advice for ever. I would also like to thank Dr. Soren Sorensen, Prof. John Campbell and Dr. Mike Dewar and Dr. Angela Sasse for their critical spirit, kindness, support and encouragement in pushing the research carriage all the way to its destination.

Thanks are also extended to the UCL (G11) networks research group, for showing me that not enough information is ever enough. In particular, I would like to thank, Piers O'Hanlon, Orion Hodson and Colin Perkins, Panos Gevros, Lambros Lambrinos, Tristan Henderson, Ian Brown, Thanassis Tirropanis, Sheng Jiang, Fulvio Risso, Dave Lewis and Kris Hasler. They have all been cherished colleagues and friend during this journey. I would also like to extend a very warm thanks to John Andrews, Denis Timm and his team, Barry Stein and his team for all their support with matters of departmental service configuration and access, hardware and system support respectively. Without their efforts this work would have encountered significant management hurdles in many ways.

I would also like to express my appreciation to Jon Crowcroft, Joe Touch and Mark Handley for shining in front of me. They have all been a guiding light and example in my research endeavours.

Special thanks go also to Carl Williams and Alper Yegin for their trust in my work, during the requirements phase for localised mobility management. Their honest comments, support, and numerous discussion on research perspective are greatly appreciated. On the research front I would like to extend my thanks to Rajiv Chakravorty, for the short but sparkling collaboration and his spirited views on research output. It has been a pleasure collaborating with him. Also thanks go to Wolfgang Fritche of IABG and James Kempf of NTTDocomo Labs, for their valuable discussions and cross-verification of Mobile IPv6 testbed performance. Special thanks are also extended to Dr. Robert Chalmers for his integrity as a researcher, implementor and individual. His objectivity set the true spin of this research effort.

No research has been ever conducted without the aid of powerful tools. To this end, I want to thank Lara Bennani and Shawn Kazzam of OPNET and Ronnie Holland of WildPackets for their kind provision of software tool licenses essential for the purposes of this research investigation. Without their kind support this research would have been delayed significantly. I would also like to thank the NS-2 research community for their silent support in completing this research endeavour.

I would like to thank Thekla Kokkali for her infinite love and understanding particularly during the final stages of the PhD write-up. It is thanks to her affection and devotion, even from a 3000-mile distance, that I was able to cope with difficult times and countless nights of work. She means everything for me.

I would like to thank Dean and Jessica Robinson, Dimitrios Miras, David Alston, Anargyros Garyfalos, Andreas Loizides, Manish Lad, Athina Markopoulou, David Lancaster, Socrates Varakliotis, Dimitrios Doumbiotis, Simon Madisson, Tasos Liolios, Rania Manousaki, Georgios Passiopoulos, Sofia Anagnostopoulou, for their kindness, support and above all, friendship during these years at good as well as difficult moments. They have all extended a precious helping hand that can never be forgotten.

Last but not least, I would like to thank *God* in whatever form or shape, for allowing me to complete this study, while still maintaining health.

While religion and science can be at times two contradicting notions, they both rely on two fundamental elements for accomplishment: *faith and persistence.* This is the single most important conclusion in this academic endeavour.

Publications

1. Williams C. (editor), Soliman H., Nordmark E., Pagtzis T., Kempf J., Malinen J., et al. "Requirements for Localised IP Mobility Management". Internet Draft, Jun. 2001. draft-ietf-mobileip-lmm-requirements-03.txt.

2. Pagtzis T., Perkins C., and Kirstein P. T. "Performance Issues for Localised IP Mobility Managment". In *Proceedings of 10th IEEE International Conference on Networks (ICON)*, Singapore, Aug. 2002.

3. Pagtzis T., Williams C., Perkins C., and Kirstein P. "Requirements for Localised IP Mobility Management". In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Louisiana, New Orleans, Mar. 2003.

4. Pagtzis T. and Kirstein P. "A Framework for Proactive Mobility in Mobile IPv6". Internet Draft, Jul. 2001. draft-pagtzis-mobileip-proactivev6-00.txt.

5. Pagtzis T. and Kirstein P.T. "Proactive Mobility for Future IP Wireless Access Networks". In *Proceedings of 6th IASTED/IEEE International Conference on Wireless and Optical Communications (WOC)*, Banff, Canada, Jul. 2002.

6. Pagtzis T., Kirstein P.T., and Hailes S. "Operational and Fairness Issues with Connection-less Traffic over IEEE802.11b". In *Proceedings of IEEE International Conference on Communications (ICC)*, pages 1905-1913, Jun. 2001.

7. Pagtzis T. and Kirstein P. T. "A Model for Proactive Seamless IP Mobility and Mobility-hop Routing". In *Proceedings of "10th IEEE International Conference on Networks (ICON)"*, Singapore, Aug. 2002.

8. Pagtzis T. and Kirstein P. T. "On the Performance of CFP Repetition Rate in IEEE 802.11 for IP traffic in Mobility-enabled WLANs". In *Proceedings of 27th Annual IEEE Conference on Local Computer Networks (LCN)*, Tampa, Florida, Nov. 2002.

9. Pagtzis T., Kirstein P. T., and Hailes S. "Proactive Seamless Mobility Management for future IP Radio Access Networks". *Journal of Computer Communications*, Special Issue on '3G Wireless and Beyond', Nov. 2003.

10. Pagtzis T., Hailes S. and Kirstein P. T. "The Case for Proactive Mobile IP". In *Proceedings of 28th Annual IEEE Conference on Local Computer Networks (LCN)*, Tampa, Florida, Oct., 2003.

11. Pagtzis T., Chakravorty R., Crowcroft J., Hailes S. and Kirstein P. T. "Proactive Mobile IPv6 for Context-aware all-IP Wireless Access Networks". In *Proceedings of Annual IEEE International Conference on Wireless Networks, Communications, and Mobile Computing Networks (WirelessComm)*, Maui, Hawaii, Oct., 2005.

Chapter 1

Introduction

1.1 Data Communications Revolution: The Internet

Over the last two decades, rapid advances in computer technology brought a new perspective in the dynamics of data processing as well as access to data resources. It is interesting to see in what ways multi-user computing paradigms carried over onto the realm of communication networks. This brings significant perspective on the shift between communication practices and its long term evolution towards ubiquitous wireless IP communications.

It all started through forms of data manipulation practices, whereby monolithic batch processing transformed into asynchronous multi-user computing systems. Users can (time-)*share* a single *processing* resource independently of their individual access pattern. The pioneering efforts of the Compatible Timesharing System (CTSS) and MAC project [1] of MIT gave rise to mainframe timesharing operating systems such as Multics [2] and its subsequent UNIX descendants. At the same time, the increasing momentum of user-centred computing emerging from these research efforts brought in the need for *access* to data resources. The computing resource was thus, required to take the form of a communication device [3].

For individual organisations, the underlying data networks supporting the computing systems, were often *private* ones, built with a technology optimised to the *specific* application [4]. At the same time, each of these computing systems was set up over experimental networks of *different* technologies. Data access was limited not only by application-specific computing systems but also to proprietary hardware and communication interfaces.

The growth in computing resources and automated data processing as a result of reductions in both size and cost of the underlying hardware transformed the design philosophy of user access to computing services. It soon became a common user requirement that a single host should be able to access *any* computing resource, irrespective of its location on the local or remote data network.

Significant contributions to such reconsideration in computing access practices were some early research experiments in the mid '70s, that solidified, under the auspices of DARPA[1], into the ARPANET project [5, 6].

There, a niche of geographically distant computing 'islands' were interconnected with the goal of providing *data communication* services to their users, by means of a universal host communication protocol; such communication interface would: (i) mask off all legacy hardware and (ii) decouple the application and computer technology from the communication mechanism. To this end, the ARPANET research effort introduced the first host-to-host protocol, called the Network Control Protocol (NCP) [7]. As a result, distributed computing and remote data access through computer interconnections were introduced [8]. The appeal of the potential from such experiments transformed what was originally a small, research computer network into a fast-growing set of network domains. A global snapshot of such network infrastructure has come to be collectively known today as the Internet.

Pivotal to the growth of Internet was the wide adoption of dynamic access allocation techniques, commonly known as packet-switching. Borrowing heavily from the paradigm of (time-)sharing a processing resource by multiple data-processing requests [9], packet-switching techniques established sharing of a *transmission* resource by multiple data-communication requests.

Instead of a continuous analogue signal over a circuit-switched connection, digital streams of binary-data blocks, known as *packet* streams[2], were created and communicated over a single transmission medium shared by more than one user. This approach signified a tremendous increase in utilisation efficiency of link connectivity (through statistical multiplexing) between networked computers; furthermore, it allowed greater flexibility in bandwidth-efficient communications between large numbers of users.

The superiority of packet-switching together with the importance of layered communication protocols in packet-networking [10] has led to the embracing of the IP protocol stack by the Internet community as the de-facto host communication protocol standard. In fact, IP has been so widely adopted that any emerging technological evo-

---

[1]Defence Advanced Research Projects Agency

[2]also referred to as packet flows

lution is conditioned by the proclivity of considering a de jure *IP-over-anything* global network, through its hardware/software implementations (Ethernet, Wireless, Cellular, ATM, xDSL).

Such a tendency is established by the ability - in retrospect, discipline - of IP to abstain from limiting assumptions about the underlying link-layer technology; it further captures the lowest but widest common denominator of interoperability and flexibility requirements in present or future data communication services. This doctrine is fostered by extensive standardisation procedures [11] in a manner that the Internet Protocol (IP) supports [12], throughout its constant evolution, a well-defined protocol suite that is reliable enough to be adopted by *any* technology.

The Internet was thus, developed to support a universal interface for *data* communications between computer hosts. Since then, however, the definition of the term 'data' has become increasingly polymorphic; its semantic expansion in terms of information content and link capacity, has enabled a variety of high-level applications services; amongst them, most popular have been electronic mail and the World Wide Web, bulk data transfer or remote login access.

Together with the evolution of the Internet, the design and philosophy of new IP protocols has realised new forms of communications. For instance, the advent of IP-Multicast has enabled a transition from the classic one-to-one (unicast) remote communication to one- or many-to-many real-time distant interactions between humans (or machines) in a network-efficient fashion.

At the application layer, new kinds of application have sprung up, focusing on *sound* and *vision*; by maintaining the temporal relationship between these two media types over computer networks, multi-sense human communication becomes a reality, incarnating Licklider's seminal vision on man-machine symbiosis [13].

The combined effect between multimedia application and multicast network deliberations provided what is called real-time multi-way communications. Of course, such applications could not be afforded without the unprecedented increase in processing power and storage capacity in (personal) computing, as well as a corresponding increase in bandwidth and network growth in the Internet.

1.2 Internet evolution: Wireless Networks

Subsequent advances in technology fomented exposure to yet another perspective of the multi-facet Internet: that of *wireless IP infrastructure networks.*

From the early days of the ARPAnet, with the ALOHA single-hop broadcast system over Satellite networks [14] and its multi-hop parallel over terrestrial packet radio (PRnet) [15], wireless networking was attempting to support user mobility, while maintaining basic IP services identical to the ones of the wireline infrastructure. Despite these efforts, the potential of wireless did not flourish; application visions propagated faster than the reality of technological advancements at the time; the *cost*, *size* and *weight* of a single PR receiver was prohibitive even for military purposes [15].

It was only during the late '80s, with the commercial deployment of GSM cellular telephony communications [16], that wireless communications received more attention, as significant reductions in both size and cost of the mobile terminal reached acceptable levels for deployment. Since then, PCS communications have been constantly evolving, offering primarily mobile telephony services, with some limited capabilities over IP.

Various other wireless technologies emerged commercially during the early '90s, embodying different physical characteristics and air interface protocols; technologies, like IEEE 802.11 WLAN [17] and HipperLAN [18], together with new generations of satellite networks were able to provide a much broader spectrum of wireless IP connectivity, each with performance characteristics driven by different communication design trade-offs.

Along with the advance of wireless technologies, significant advances in power control [19] and battery cell technology [20] have made it possible for mobile terminals to stay attached with a network for several hours[3] with a single battery charge. Continual improvements in chip design, miniaturisation, and integration have made it possible to add increasing functionality to small user devices; new short-range wireless technologies like Bluetooth [21] and Zigbee [22] address wireless interconnect in an increasingly flexible way. Smart phones and multi function PDAs [23] have emerged in the market to provide combined computing and communication capabilities. The dumb, cheap mobile terminal is gradually replaced by sophisticated mobile handsets with common features such as voice recognition, location tracking and touch screens. Audio is already the norm in these devices, while miniature cameras are introduced as integral part of the communication interface.

---

[3]depending on the pattern of use

1.3 Task performance and novel IP applications

Application development, on the other hand, tracks advances in hardware features by realising multimedia services in the most unassuming ways; novel audio-visual applications appear to take pole position in daily human tasks, such as navigation or mobile collaborative group work in fields like medicine. These services make more sense for the mobile user if supported on-demand everywhere, instead of resorting to tentative static accumulation of data, in view of future potential need.

Traditional multimedia applications are faced, however, with important performance obstacles against the dynamic nature of the wireless medium, since by-design they considered wireline network deployment. The real-time requirement for multimedia applications imposes naturally stringent delay constraints; large delays translate to significant packet loss, while large bursts of packet loss disrupt the communication pattern between peers, rendering communication interactions unintelligible [24]; variation in arrival delay of communicated traffic cannot exceed a few hundreds of milliseconds [25].

Wireless communications require some degree of stream survivability on the part of the sender, the receiver or the network itself; these need to be considered by the network and/or the application layer. Transmission resiliency becomes an essential function in counteracting the effects of loss over the wireless interface [26, 27]. This puts even more strain on the network/application before the mobile device can utilise the application's full capabilities.

Novel types of data dissemination also become important. Wireless hand-held devices and users integrate interactive communications or information retrieval as a integral part of their main task activity (driving, operating, pursuing, rescue, defending, etc). Interactive communications must maintain intelligible levels if task-specific performance is to be sustained. Both the network and the application layer must constantly assess delay-prone packet deliberations, to preserve adherence to such guarantees in a best-effort inter-network.

1.4 Wireless Internet: a maturing state of affairs

Wireless computing technologies have reached a stage of maturity that begin to realise practical forms of truly open wireless access capabilities; some of them are: global coverage, *'always connected'*, uniform performance characteristics, *a priori* mobility patterns, seamlessness, or connection transparency while in communications. Mobile users envisage (and expect) a ubiquitous convergence of access onto the wireline network infrastructure, through a heterogeneous wireless last-hop.

These requirements engender a paradigm shift in the traditional access practices of fixed computer networks, collectively known as mobile networking. Under these new access practices, users of portable wireless computing devices require access to some shared wireless network infrastructure; this is independent of their physical location, typically, while in transit.

Hence, the evolution user-centred network access practices receives now the form of ubiquitous continuity in mobile access of the Internet infrastructure. This is particularly important at a time when physical mobility is encouraged and supported at a global scale, independent of national or international geographic borders.

1.5 Wireless link challenges: An evolving maturity plan

Despite the advances in wireless technologies (see Annex B.1), there exists *no* single wireless interface that provides low latency, high bandwidth and wide area data services, to a large number of mobile users.

In addition, wireless networking, offers certain advantages at the expense of some quite constraining drawbacks. For instance, from the perspective of medium access, a wireless station cannot receive and transmit over the same carrier channel simultaneously, in contrast to its wireline counterpart; this is because wireless modems are predominantly half duplex, allowing only receive or transmit at a single time instant, over the same channel.

Due to different mobility patterns, terrain obstructions, or even weather conditions affecting the air interface [28], the mobile terminal is guaranteed to experience different signal attenuation during reception or transmission. As such, the behaviour of the wireless broadcast channel is different during *any* two moments where transmission or reception may take place; a wireless link *cannot*, in general, guarantee the same channel quality within all locations of coverage area.

In addition, propagation effects such as multi-path fading or path loss [29] can impede or err the arriving bit-stream at the receiver. The manifest fragility of a wireless link within harsh propagation environments transforms into increased bit error rate (BER) and associated packet loss. As a result, packet loss over a wireless link is associated with wireless link *corruption* rather than routing path congestion.

A single IP subnet, made available over the air interface of two or more wireless *access points* (APs), presents the task of sustaining its availability over any of these APs in a transparent manner; that is, the wireless host must be able to sustain continuous subnet connectivity, through transparent association with each of the APs as it transits over its coverage area. To provide such capabilities, the link-layer of wireless technologies such as IEEE 802.11 ensure transparent access to the host over a single IP network, irrespective of its underlying AP association.

1.5.1 Wireless IP hosts on the move

Besides the issues, arising as a result of the host's physical mobility, significant application performance challenges emerge from the actual *detachment/attachment* of the wireless host from the associated IP network, as a result of its *IP network* mobility.

In addition to physical or link-layer mobility, the need for IP network mobility arises, when the wireless host pursues roves beyond the coverage boundaries of the existing point of IP attachment.

However, IP networks were not designed with IP mobility in mind. An IP application associates explicitly[4] its connection with the host's IP address, when establishing a communication path with an identical application instance residing in the remote host. This is because in this connection the IP address *identifies* each communication party. As a result, the identity of both communication parties must be preserved irrespective of the user/host physical movement. However, under traditional Internet semantics, an IP address serves a dual role: (i) identification and (ii) location routing.

Physical movement of the host extending beyond the wireless coverage of one IP subnet causes the host to alter its location within the network topology. In such cases, the host must attain a new IP address if it is to remain globally reachable[5] as well as locally identifiable by the new IP network segment. However, a change of IP address implies a change of the host's identity; as such, the application layer of the IP-mobile host must reset and re-establish any existing connections.

As a result, both network and application layers of the wireless host, are faced with severe dis-connectivity when the host moves between IP networks. *IP Mobility Management* becomes, thus, of importance for hosts that need to move beyond the coverage boundaries of a single wireless IP subnet.

This challenge effectively becomes one of path rerouting that must be managed at the network layer, if applications and their underlying transport are to avoid dis-

[4]through the opening of a socket the host binds to the IP address and a relevant port.

[5]reachability implies that the host can send/receive traffic as long as its peers can identify the host and connect to it.

connectivity; routes in the communication path between the host and its peers must change. At the same time mobility management at the network layer must remain independent of assumptions about the underlying wireless technology, if it is to ensure that the supporting protocols can be integrated across any wireless platform. To this end, a number of IP mobility management mechanisms have been proposed or standardised [30, 31, 32, 33]. Amongst them, Mobile IP [31, 32] arises as the de facto standard for IP mobility management, providing identification and routing transparency in the IP mobile deliberations of wireless host.

From a network provisioning perspective, IP mobility management is expected to bring a tighter convergence relationship between networks and their utility, facilitating task performance. It enables the Internet to act as an IP-transparent communication infrastructure. However, such transparency does not extend to aspects of delay performance critical for interactive IP application services.

1.6 Interactive Multimedia Applications

Currently, the Internet, in its *IP-stationary* incarnation, serves a large part of the user's daily task activities that are generally insensitive to delay: from email correspondence and information retrieval, to streamed audio/visual entertainment or commercial and accounting transactions. In addition, access to the Internet and most of its applications are provided as *best-effort* services. These are commodified through flat-rate subscription models for network access, free application usage, and high user tolerance to lower application service quality.

Delay-inelastic application services have only recently started making their tenuous appearance as a tariff-based application service, through IP-stationary communication networks. Amongst them, IP telephony is an application with enormous market potential, served until very recently by the traditional public telecommunication (PSTN) carriers. Video conferencing is also becoming increasingly popular for business meeting communications with participants in remote sites, tele-presence in educational or medical environments, or even live broadcast news coverage from remote locations.

The above indicate that as the Internet evolves to a commercial network, new business models arise and fees may be associated with novel application services or communication requirements. From a user perspective, IP-mobile extensions to wireless IP connectivity and geographical reach make the use of commercial IP networks even more attractive. Nonetheless, tariff-based application services typically raise user expectations with respect to service quality.

Therefore, there is a long-term motivation for creating an efficient IP mobile network capable of supporting successfully *delay-inelastic* application services.

1.6.1 Requirements

Interactive IP applications are such a type of service with higher performance requirements than the traditional best-effort data applications. Real-time multimedia applications require low latency [34] and reasonably good quality [35, 36], at a level similar to public telephone networks.

For the wireless Internet to constitute an attractive alternative to the traditional wireline networks, or to operate in conjunction with them, it is critical that it meets the above requirements at similar or better quality levels.

IP Telephony

In the context of IP networks, telephony services, are known as Voice over IP (VoIP). A VoIP flow depending on the encoding employed, such as G.711, G.723.1 or G.729, can generate data rates between 5.33 and 64 Kbps [37, 38, 39, 40]. Packets are generated isochronously at the supported packetization rate through the real time protocol (RTP) containing a fixed size payload.

To maintain a conversation at good quality levels, a VoIP flow requires low packet loss rates. Loss rates up to 10% may be tolerated depending on the type of packet concealment technique employed by the decoder on the side of the receiver [26, 27].

To sustain intelligibility of VoIP communications the total end-to-end delay should remain below 150ms or lower, for highly interactive conversations [41]. Delays in the range of 150-400ms are considered acceptable, although the annoyance becomes perceptible; delays greater than 400ms are considered intolerable and thus unacceptable for effective communication.

In summary, interactive multimedia applications need: (i) little or no loss for good speech and video quality (ii) low delay for interactive communication and (iii) low or no delay variability for continuous play-out. It is questionable, how such requirements can be supported in an IP-mobile wireless Internet, since currently the Internet cannot provide sufficient guarantees for their performance.

1.7 The need for Seamless IP Mobility Management

Cellular networks in the past decade have been built primarily to deliver voice services. This class of mobile networks has untethered voice telephony while coupling it with user mobility. Due to its immense popularity the number of mobile subscribers is set to overtake land-line subscription levels in the next year or two. Perhaps the only network that has parallelled (and exceeded) the growth of cellular networks and mobile subscribers in the '90s is the Internet.

If the paradigm set by cellular networks is to apply successfully over a wireless IP-mobile network infrastructure, it becomes clear that efficient forms of *delay-transparent* IP mobility management *are essential* to support successfully interactive IP application services. This thesis advocates the importance of future IP mobility management system architectures supporting robustly the performance delay requirements of interactive multimedia applications, at the network layer.

Support of interactive IP applications at the network layer entails the provision of service performance consistent with human perception or user expectation. In IP wireless access networks, such challenge is amplified by yet another significant factor pertaining to IP mobility: the management of service disruption as a result of host transitions between different networks. These transitions are known as *IP handoffs*, emerging on the node's wireless transit path.

Initial research efforts towards the introduction of IP mobility management into the network stack of the IPv4 protocol family, have shown that an IPv4 handoff can introduce delays of a multiple seconds in the end-to-end delay of the host's received traffic [42]. Such delays have been unacceptable not only for interactive applications, but also for delay-elastic applications [43].

With the introduction of next generation Internet Protocol version 6 (IPv6), a number of components from standard IP(v4) mobility management have been combined into a recent protocol standard, known as Mobile IPv6. Despite such enhancements, however, Mobile IPv6 appears to introduce excessive latency during an IPv6 handoff for the purposes of interactive IP applications.

The aim of this thesis is to investigate factors that can impede delay-transparent performance in interactive applications, over the Mobile IPv6 management standard. Subsequent parts of this study, reconsider the architectural framework set by the Mobile IPv6 specification and its derivatives. To this end, this thesis looks into efficient alternative mechanisms in support of delay-seamlessness in IPv6 mobility management, for interactive IP application services.

1.8 Thesis Scope

Investigations in this research effort are concerned with:

- state-of-art IP mobility management mechanisms. Identifying their strengths and associated design trade-offs or emergent shortcomings would enable further investigations to be based on a solid set of requirements for robust delay-seamless IPv6 mobility management.

- the delay performance shortcomings of Mobile IPv6 during the mobile node's (MN) IPv6 handoff. These are identified and validated through experimental measurements over a real network implementation.

- proposing an IPv6 mobility management (MM) architecture that addresses the deficiency of re-connection delay transparency in the face of multi-context state establishment that sustains the IP connectivity of the mobile host.

- evaluating the aforementioned architecture by means of discrete event simulations. This is performed through two core functions tracking the performance of the proposed architecture: (i) handoff (ii) flow forwarding management.

Handoff-management performance is pursued by demonstrating delay seamlessness over addressing and routing, collectively identified as *IP Roaming state.* This is the minimum state establishment requirement, in the context of MN's IP connectivity, that is necessary and sufficient to establish the case of advanced IPv6 mobility management mechanisms for the purposes of delay-seamless performance.

Discrete event simulation is an essential part in system design since it allows efficient exploration of the parameter space. Such investigation encompasses operational scenarios that are cost-inefficient and time-intensive to achieve through real-world experimental testbed configurations. This is the case for wireless IP-mobile networks, since a significant number of MNs is required to participate in a single experiment investigating only a *limited* range of the required parameter space. Despite the appeal of a real network configuration of the sort, such investigation effort provides only limited insight into the dynamics of the system under investigation.

Our research investigations are scoped for IP protocol version 6 (IPv6) with particular focus over IEEE 802.11 wireless LANs. Reason for such focus is the excessive measure of delay impose by the respective link-layer handoff function, in comparison to other wireless technologies (e.g. cellular).

In principle, the proposed IP mobility management architecture is designed to remain independent of assumptions about the underlying wireless technology. This is because IP handoff performance is assessed through statistical compliance over specific delay bounds. The extent to which such compliance can be achieved is dependent on the delay performance of the underlying link-layer handoff. As shown in Chapter 3 link-layer handoff delay is an integral component of the total IP handoff delay experience by the MN.

Despite the above dependence, the handoff management function of the proposed architecture focuses on latency incurred at the *network* layer. Hence the performance evaluation of the proposed handoff management mechanism remains *independent* of the underlying technology. Assuming native IPv6 signalling[6], the proposed handoff management function can be, thus, generalised over *any* wireless technology.

Link-layer (L2) handoff performance becomes of importance, however, to the proposed function of flow forwarding management. To this end, we focus on the L2-handoff delay profile of IEEE 802.11, since GPRS/UMTS L2-handoff delay performance appears *not* to affect significantly the performance of MN's IPv6 handoff.

We evaluate the proposed IPv6 mobility management architecture against Mobile IPv6 as the standard in IP mobility management.

We further compare and analyse the protocol operations of the proposed mobility management architecture with Fast Handoffs in Mobile IPv6 (FMIPv6), as the emerging solution for delay-transparent mobility management.

Security issues pertaining to the robustness of the proposed mobility management model are evaluated architecturally.

1.9 Research Contributions

This thesis advocates **the need for delay-transparent IPv6 mobility management to support robustly interactive IP application services over wireless access networks**.

To this end, this research explores the hypothesis that *reactive registration with the new point of attachment (PoA) in current IPv6 mobility management standards is insufficient to address delay transparency; by means of proactive registration to candidate PoAs, the IP handoff management task can support realistically transmission delay seamlessness.* Registration is defined as the process of state establishment of one (or

[6]The 3GPP forum working on UMTS technologies is currently aligning towards such mode of IP signalling.

more) contexts pertaining to the continuous IP connectivity of the mobile node.

Furthermore, the proposed thesis argues that *reliance on the completion of the IP handoff to resume packet transmissions between the MN and its peers, as effected by reactive IP mobility management signalling, is insufficient to support transmission delay seamlessness during IP handoffs; by means of proactive flow forwarding management, such reliance can be eliminated and transmission delay seamlessness can be preserved.*

It is important to emphasise the difference between the issue of handoff and flow forwarding management; handoff management deals with the measure of delay incurred during an IP handoff. Flow forwarding management focuses on sustaining packet delivery towards the MN, during an IP handoff.

To this end, our research investigations will be making the following contributions:

1. an in-depth review of the state-of-art in mobility management with a critical perspective on strengths and associated design trade-offs of alternative mobility management proposals.

2. a detailed experimental insight on practical performance shortcomings of the Mobile IPv6 standard, with respect to interactive real-time IP application services. We reveal performance issues related to the operation of core IPv6 protocols such as Neighbour Discovery and how these impact the observed measure of IPv6 handoff delay.

3. a novel IPv6 Mobility Management (IPMM) architecture establishing and preserving the seamlessness principle, in terms of transmission delay, during IP handoffs realised through multi-context state establishment. Such state is typically part of the IP connectivity state of the MN. This encompasses the design and evaluation of a novel IP handoff management approach that practically eliminates: (i) network-layer handoff delay (ii) dependence on core IPv6 control signalling.

4. as part of the proposed IPMM architecture, a novel protocol mechanism for the discovery of candidate points of attachment (PoA) for the purposes of MN's IPv6 handoff, tightly coupled to the handoff management function. The importance of such contribution pertains to the extensions of the protocol's semantics onto: (i) context-specific state *establishment/relocation*, (ii) identification and exchange of *capabilities* of candidate points of attachment. The latter is employed by the MN as an indicator of IP handoff diversity for the purposes of maximising MN's service utility.

5. an efficient form of flow forwarding management, alleviating service disruption of active IP transmissions between the MN and its peers, during an IP handoff. With emphasis on delay-prone wireless technologies, the proposed flow forwarding mechanism is enhanced by an optimised measure of L2-handoff latency over IEEE 802.11 WLAN networks.

1.10 Thesis Structure

This thesis is structured as follows. Chapter 2 presents an in-depth review of the state-of-art in IP mobility management. It engages in a critical discussion on the issues arising as a result of the evolution of the IP mobility management function in the face of novel application services and performance of existing mobility management standards. Related issues in this chapter have also been presented in [44], [45, 46].

Chapter 3 investigates the performance of the dominant IP mobility management standard, Mobile IPv6, through experimental measurements. We devise a detailed wireless experimental testbed, MobiRigv6, deploying interactive IP applications over IPv6. Experimental investigations focus on IPv6 telephony, as the interactive IP application of choice, effected over infrastructure IEEE 802.11 WLANs. To provide an accurate account of IPv6 mobility performance, we isolate the handoff performance of the link-layer from the performance of the network-layer from experimental traces, and report individually on each of the two. Related issues in this chapter have also been presented in [45], [47] and [48].

We derive performance measures of quantitative metrics such as handoff delay, associated packet loss and jitter. These measures establish the performance of a widely accepted IP mobility management mechanism as the control hypothesis utilised for comparative purposes in subsequent investigations.

Chapter 4 presents a novel IP mobility management architecture identified as Proactive Mobile IPv6 (PMIPv6), that promotes the notion of in-advance control deliberations in view of MN's next IPv6 handoff, at the network layer.

The chapter first presents a dynamic handoff AR discovery mechanism. Based on the notion of access *Routing/Mobility Neighbourhoods* within a wireless access infrastructure, ARs *'conspire'* to provide essential information pertaining to MN's next IP handoff, in advance of its next IP transition. To this end, we identify a proactive state establishment protocol to support state identification and generation pertinent to MN's next IP handoff. The notion of state expands onto abstractions of metric-based capabilities as a means of allowing PoA selectivity and intelligent handoff control by the MN in an effort to maximise its measure of service utility.

Subsequent parts of this chapter investigate the performance of Handoff AR Discovery (HARD) and its impact on MN's next IP handoff delay. We assess key factors characterising the non-determinism of MN's mobility pattern, such as speed and pause period by means of simulations. Furthermore, the impact of density in points of attachment (PoA) and MN population on the convergence of HARD state is evaluated. Parts of this work have also been presented in [49] and [50].

The last part of this chapter evaluates the relative benefit of handoff selectivity as an approach of enabling the MN to select its PoA over its next IP handoff transition.

Chapter 5 presents HandoffCast, a novel IP mobility management mechanism that supports delay-transparent forwarding of MN's flows towards its new point of attachment, *during* an IP handoff. HandoffCast is an integral component of the IP mobility management architecture proposed in Chapter 4.

To complement the performance of HandoffCast over 802.11 technologies, we devise and evaluate a simple optimisation to reduce the measure of persistent handoff delay emerging as a result of long link-layer handoff delays over WLAN networks. Part of the 802.11 link-layer investigation is presented in [51].

Chapter 6 presents a comparative performance analysis between the proposed mobility management architecture and Fast Handoffs for Mobile IPv6 (FMIPv6); we provide qualitative and quantitative evidence demonstrating significantly better performance in the proposed mobility management system over the FMIPv6 counterpart, in the face of erroneous tunnel setup as a result of ping-pong effects. Parts of this work appear in [52].

Chapter 7 summarises the contributions of this thesis and reviews possible next steps of future research work.

Ultimately, Annexes B-G present auxiliary information and additional experimental results supplementing the completeness and validity of our contributions.

In particular, Annex B presents a brief overview of new and emerging wireless technologies, as well as popular wireless technology alternatives.

Annex C presents a critical review of mobility management proposal and their signalling performance, covering the complete spectrum of macro and micro-mobility management protocols elaborated in Chapter 2.

Annex D presents additional results pertaining to MIPv6 handoff performance (analysed in Chapter 3), analysed from extended trace measurements.

Annex E presents results pertaining to the performance of Proactive handoff management (analysed in Chapter 4), derived through discrete event simulations.

Annex F presents additional performance results from the simulation analysis of the proposed HandoffCast flow forwarding management, detailed in Chapter 5.

Annex G provides the analytic representation of a three-cell overlap model employed in the analysis of Chapter 6.

Annex A provides a glossary of frequently used terms or abbreviations comprising part of the IP mobility management engineering parlance.

Chapter 2

A survey of IP Mobility Management mechanisms

Understanding the different classes of IP mobility management architectures, protocols or functions is important to consider, assess and evaluate the respective functional components and their associated performance trade-offs. This chapter provides a critical survey of the state-of-art in IP mobility management solutions, encompassing both protocols and architectures.

The chapter commences with a brief overview of IP Mobility management principles originating from pioneering work on host mobility and cellular systems and shaped by dominant mobility management standards. Utilising the former as the focal point, new issues and challenges are then introduced, pertaining to the evolution of IP mobility management requirements with specific emphasis to support of interactive real-time services.

We then proceed to review the current state of the art in IP mobility management mechanisms, focusing on micro-mobility and macro-mobility protocols. This review is augmented by a critical assessment of the techniques adopted by these mechanisms. The latter serves as a taxonomic substrate of techniques associated with critical functions of IP mobility management, together with their respective strengths and weaknesses. To this end, we focus on design choices and performance trade-offs associated with dominant or emerging IP mobility management schemes proposed to-date. The latter is subsequently feeding into requirements analysis conducted in later parts of this thesis.

2.1 Introduction

The advances of wireless technologies [53, 54, 55, 56, 57] and portable computing terminals [58, 59, 60] are reaching a state of maturity, where users envisage (and expect) a convergence in the wireless/wired network infrastructure that allow new diverse access capabilities such as: access *on the move*, global span, **always connected**, uniform performance characteristics, seamlessness, IP transparency, ad hoc connectivity. These capabilities engender a paradigm shift in the traditional access practices of fixed computer networks, collectively known as *mobile networking*. Under these new access practices, users in command of portable wireless computing devices require access to some packet-switched, **all-IP**, wireless network infrastructure independent of their physical location, while in transit.

Furthermore, the notion of ubiquitous computing [61] as enabled over mobile networking practices has opened up new possibilities for novel kinds of multimedia applications on the move: navigation [62, 63, 64], personal locator services [65] interactive audio/video [66, 57], network games [67]. The capabilities of today's wireless terminals [68] have been constantly increasing such that processor-intensive services, like interactive multimedia, are becoming the expected norm rather than extreme futuristic visions.

Real-time dissemination of multimedia information becomes now even more important as mobile devices and users integrate information retrieval as a peripheral task of their main activity (driving, operating, walking, acting, etc). These activities require low latencies if communications are to sustain real-time guarantees in terms of both task performance and communicated information.

Ubiquity has introduced further the potential for *nomadic* communications [69, 70]. From a role-mobility perspective, mobile users require the appropriate network/application level support to assist them transparently, throughout their various role-tasks entailing network applications, multimedia in particular, while moving to their ignorance over multiple coverage areas that span geographically towards some destination.

With a ubiquitous wireless Internet in mind, it becomes apparent that IP attains a second role besides acting as an application layer unifier; it allows a sustainable access convergence through an evolving network infrastructure supporting multiple wireless technologies, both at link (L2) or physical (L1) layers. From the arising multiplicity of wireless technologies, the emergent *duality* in the use of IP has a significant impact on network efficiency and access performance, if it is to sustain continuity in the communication practices of the mobile user.

At the same time, various industry consortia, such as WiFi [71] and 3GPP/3GPP2

[72, 73] perceive network access architectures from different, at times conflicting, service standpoints. In this light, reconciliating access architectures that support packet-switched best effort access, while providing assured circuit-switched type of service quality, presents its own performance challenges.

The aim of the wireless evolution remains, none the less, common: ubiquitous availability of IP-based multimedia application services. For that purpose it encompasses adoption of an IP-based transport as well as the integration of Internet Engineering Task Force (IETF) protocols for *key* functions such as wide-area mobility support, signalling, access control and billing, or quality of service. It is thus, becoming increasingly popular to call any network that encompasses the aforementioned components in its network access architecture, an *all-IP* network.

2.1.1 Mobility Support

The departure from the traditional circuit-switched model of personal communication systems (PCS), towards all-IP wireless networks places particular emphasis on efficient mobility management mechanisms. This is so, because mobility management is a 'cornerstone' function that *implements* as well as *preserves* continuity of communications between the wireless host on the move and its peers.

The former statements purposely avoid the immediate qualification of the 'mobility management' task as solely 'IP-based'. This is because, in any wireless Internet Service Provisioning (ISP) domain, the underlying network infrastructure, encompasses *two* core tiers of device connectivity. These tiers are:

- (Wireless) Access Point (AP): also referred to as *Base Stations*, depending on the underlying wireless technology. Such device implements control functions at the link and physical layer, allowing communications of the wireless host at the physical (L1) and link (L2) layers. At L1 the wireless interface implements *modulation schemes* [74, 75, 76], that prescribe the encoding as well as transmission/reception timing of a modulated bit-stream, propagated over the air interface. At L2 the wireless interface implements *coordinated access control* [77, 78, 79] of the wireless medium (MAC). The MAC layer (L2) effectively imposes a host ordering on link access, to prevent collisions amongst wireless hosts contending for transmission access.

- Access Routers (AR) : Such device implements a forwarding and a routing function allowing to route packets amongst *multiple* network interfaces towards their destination.

In infrastructure networks an AP extends the domain's AR last-hop over the air interface of the wireless technology at hand. In this manner, one or more wireless hosts *associate* or *attach* with an AP to access the associated IP network domain. Hence, in infrastructure networks, APs are also knows as point to multi-point 'bridges'.

One or more APs may be connected to the *same* network interface of an AR; access to any of these APs implies access to the *same* IP subnetwork. The sole difference between one and many AP(s) connecting to a *single* AR interface, is the increase in wireless bandwidth and coverage area as shown in figures 2.1 (A).

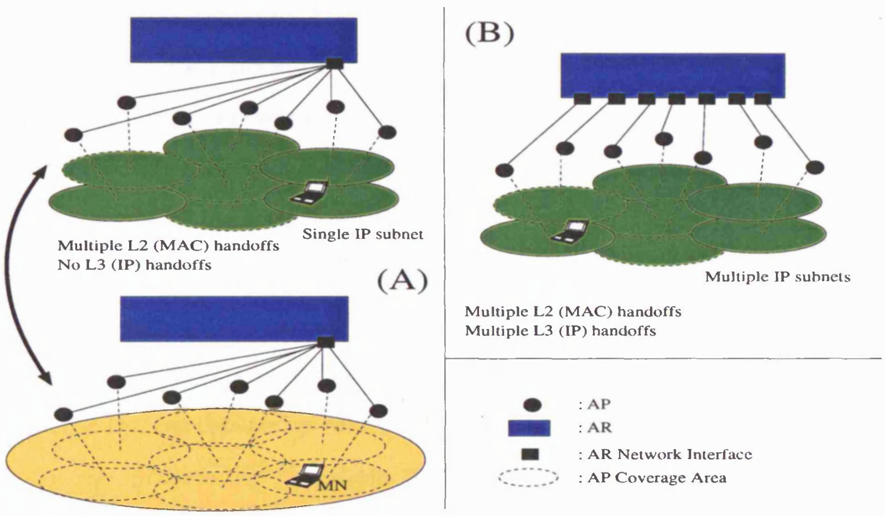


Figure 2.1: Many-to-One and One-to-One network configuration between APs and AR interfaces and its implication on L2 and IP handoffs

Alternatively multiple APs may be connected each to *different* network interfaces of an AR. In this manner, an AR implements multiple *different* IP subnetworks manifested as *IP cells* serviced by individual APs, as shown in figure 2.1 (B).

Transition of the wireless host between different APs under the *same* IP subnet, incurs *multiple* link-layer (L2) handoffs, but *no* IP handoff. On the contrary, transition of the host between different APs under *different* IP subnets incurs multiple L2 handoffs and multiple network layer (L3) handoffs. The system of AP-AR is also identified for mobility management purposes as point of (IP) attachment (PoA); this is independent of the underlying wireless technology, be it 802.11b/a [80, 17], GPRS [81], Bluetooth [21] or UMTS [73, 72]. Hence, *host mobility* may be classified as:

- link-layer or L2 mobility: it effects the physical transition of the host's network interface between homogeneous links, by employing physical medium characteristics and link-layer control.

- network-layer or *IP mobility* : it abstracts the underlying link-layer by effecting transition between different IP networks. IP Mobility follows *after*, but remains independent of link-layer mobility and hence makes no assumptions about the underlying L2 technology.

In the context of different wireless technologies the semantics of L2 and IP mobility attains a different scale of applicability as a function of the coverage area of the AP. Chiussi et al [82], abstracts the notion of L2 and IP mobility to *access* and *wide-area* mobility to accommodate the semantics of IP mobility management onto third-generation (3G) cellular networks, as a result of the large coverage areas afforded by the AP tier of cellular networks. In such networks, IP handoffs are less frequent in relation to the ME[1] velocity, since large geographical areas are covered by AP-clusters comprising a single IP subnet.

It can be seen that the form of L2 mobility is dependent on the wireless technology at hand. For instance, the link-layer of cellular networks effects L2 handoffs by means of dedicated link-layer functions such as *angle of arrival estimation* of the pilot signal [83, 84] and/or by coordinated power control [85] among APs[2] to guarantee statistically a certain access probability threshold or L2 handoff delay bound. On the contrary, IEEE802.11 networks, commonly referred to as WLANs, effect L2 handoffs by plain control of SNR thresholds [86].

From this perspective, *different wireless technologies impose different performance constraints* to the implementation of mobility support for a given wireless IP network domain. Hence, performance of a wireless link-layer (L2) with respect to L2 mobility, as well as its impact on IP mobility performance, is *not* generalisable across all wireless technologies. This important to acknowledge since an exhaustive investigation on the effects of wireless technologies onto IP mobility is out of the scope of this study.

For this reason, we do not engage in a *generic* presentation of link-layer handoffs since the former is strictly technology-dependent; instead, we focus elaboration on the main principle of mobility management at the network layer. Any explicit mention to L2 handoff performance is provided *where necessary* for the purposes of indicating

[1]Mobile Equipment

[2]To aid elaboration, the terms Access Point (AP) and Base Station (BS) are considered to be equivalent and thus, used interchangeably.

factors of influence manifested by the wireless medium over IPv6 mobility management.

The following section presents the main principles of IP mobility management.

2.2 Overview of IP Mobility Management principles

2.2.1 Mobility Management Predecessors: Cellular Networks

The mobile communications paradigm set by *circuit-switched* cellular networks, such as GSM [16], has paved the architectural path for IP network mobility. This has been done by lending - or more accurately porting - in one form or another, its underlying architectural configuration onto *packet-switched* IP networks. For this reason, a brief description of their vital management functions is essential to bring subsequent IP mobility management techniques or proposals into perspective. We refrain from an in-depth description of the GSM protocol or its functions as it expands beyond the scope of IP mobility management.

Core Architectural Paradigm

Cellular networks, comprise of three fundamental management tasks: (i) *handoff* (ii) *location* of the *mobile terminal equipment* (MTE) onto the cellular infrastructure (iii) *routing* of voice calls to the MTE. The process of handoff is typically managed by the 'link-layer' of the cellular network with no intervention by location management or call routing. Hence, for location management and call routing purposes, handoff management is a transparent function the internals of which can be ignored. Suffice it to say that the results of handoff management trigger the appropriate location management functions; the latter subsequently enables call routing to the MTE, irrespective of its whereabouts within the cellular network domain. What follows is an extremely simplified view of the mobility management operations within a cellular network that aid in understanding subsequent IP mobility management proposals and their origins.

Cellular location management involves the cooperation of a number of *registers* within what is identified as *home* and *visited* network. Such networks comprise the greater cellular domain infrastructure, aiming to provide ubiquitous cellular telephony services to mobile subscribers. A *home location register* (HLR) residing at the respective cellular network, maintains a user profile for each MTE subscriber. Within a visited network the respective *visited location register* (VLR) maintains the current cell attachment of the MTE.

As the MTE roves within the cellular infrastructure, comprising of multiple visited networks, the VLR tracks MTE's location, with infrequent location updates to the

MTE's HLR. In this manner, the HLR maintains an accurate pointer of the MTE's location within the visited network. It is important to note, that updates within one (or more) location area(s) (LA) managed by the VLR, are processed at the local VLR and do *not* require signalling with the HLR. Hence, the HLR refers always to a VLR *before* reaching the MTE.

To further limit signalling within the visited network, location updates are not necessarily sent for each cellular handoff; this is because the MTE may be *idle* while moving[3]. Instead, *paging* is used to pinpoint MTE's actual cell location within the LA, at call-setup time; this allows MTE's to conserve power.

While appealing, paging incurs certain signalling trade-offs over maintaining MTE's current location. While frequent updates consume precious bandwidth and energy, paging over large areas (one or more LAs) results in bandwidth wastage[4] for all base-stations other than the one accommodating the MTE. From a signalling perspective, while optimisations for such design trade-offs have been proposed [87, 88], it becomes apparent that for a decreasing cell size, paging and subsequently location management becomes more expensive. Such trade-offs have similar performance repercussions when similar architectures are adopted over packet switched networks as seen in subsequent sections.

With the above architectural paradigm in mind, we embark on an in-depth investigation of IP mobility management techniques, their evolutions as well as issues and challenges emerging from their application. This would aid understanding of our investigations in subsequent chapters.

2.2.2 Host mobility and limitations of traditional networking

An IP address acts both as a *location/routing* as well as *host identifier* that tracks its associated host over a particular network link. Under conventional networking practices, when the (wireless) host migrates to a different IP subnet, namely, perform an *IP handoff*, its IP address *must* change to one that is topologically correct over the new link. The immediate limitation arising from such movement is that the host cannot be located anymore by means of its original, commonly known as *home*, address [89]. For peers that attempt to contact the mobile host by means of its IP address or DNS name, the host appears to be unreachable; the mobile host appears unable to inform its peers of its new IP address since it bears no knowledge that it has been contacted by some

[3]which is the typical case, otherwise subscriber billing and congestion would grow prohibitively

[4]In GSM networks bandwidth is very small and hence, from a revenue-oriented pespective scarce and precious to be wasted in signalling.

communicating peer.

Beyond disruption of reachability, the mobile host experiences a much more significant effect that acts to the detriment of any active communications with peers. Typically every active network connection of a host, in the form of *sockets*, binds *explicitly* to its existing IP address and port. The traditional form of network connection establishment imposes the limitation that acquisition of a new IP address, causes all connections maintained by a host through sockets, to be dropped and re-established. Thus, even if the host can attain a new IP address instantly, it will have to drop all of its existing connections and, upon reconfiguration, re-establish them.

As a result of the aforementioned limitations, physical mobility for the IP host remains limited over a single wireless IP link, if it is to afford reachability or sustain active communications with peer hosts.

2.2.3 The birth of IP macro-mobility

To remove such limitation, mobility extensions over IP were first proposed by Ioannidis et al in [89], known as Mobile IP Control Protocol (MICP). Under this proposal, Mobile Support Station (MSS) gateways support the mobility of a host acting both as points of IP attachment (PoAs) associated with the mobile host (MH), as well as location directories each for the MSS's own set of associated MHs. Communication between a peer and the MH is effected by means of *tunnels* between MSSs. For this purpose, a simple paging function is supported through multicast, to identify the MSS providing reachability to the required MH, in cases where the location of the MH was not known.

It is important to note that in MICP the MH retained only its home IP address; the MSS handles special ARP caching for forwarding purposes; for this purpose traffic arrives encapsulated at the MSS, where it was decapsulated and subsequently forwarded. Subsequent work in [90] extended MICP with minor optimisations to support an intra/inter-campus mobile inter-networking architecture.

Subsequent IP mobility protocols proposals, such as Loose Source Routing [91], IMHP [92], its derivative, Mobile IP [31], forwarding-pointer caching at ARs [93] and VIP [94], *simplified* routing and location management overheads, each by means of different signalling techniques supporting variant performance and trade-offs. For instance, in [91] the loose source routing option was included on packets sent by the MN to CN, to include MN's *Foreign Agent*[5] (FA) address; in this manner on its response

[5] A foreign agent is essentially a mobility management function that may be hosted onto an access router (AR) device. In this context an FA is equivalent to a mobility-aware AR

the CN can access the MN through its associated FA.

A common element in all the above proposals, is the usage of *two* IP addresses by the mobile node (MN). However, with the exception of Mobile IP, in all the above systems, source routes and location updates were not authenticated, providing an opening for arbitrary redirect attacks, due to unauthenticated location updates.

Routing optimisations with respect to IP mobility were further proposed by [95], whereby a secure non-triangular routing approach utilised standard unicast routing for Mobile IP; alternative optimisation were proposed in MINT [96], Mobility Support Network [97], and Mobile Inter-networking Architecture [90].

For instance, VIP proposed a variant of the generic mobility architecture subsequently proposed by Bhagwat [98] whereby intermediate routers can cache location update and act as address translation agents. Such approach however, incurs increased signalling overheads for updating address mappings, given that entries on intermediate routers are flushed through directed broadcasts. A comprehensive comparison of several of the above mobility management approaches is available in [99].

2.2.4 Establishment of IP macro-mobility standards

Of all the above protocols, Mobile IP [31] has become the most popular; it is simple, compatible with existing applications and hosts and places no special burdens on normal IP routers. Chesire and Baker [100] emphasise that the Mobile IPv4 protocol should not be a routing panacea; it should embrace flexibility, such that optimisations (including Mobile IP itself as a value-added network service) are applicable only where appropriate under different network conditions.

Nonetheless, Mobile IP remains the most widely accepted macro-mobility protocol; its importance is such that the protocol merits a short description of its core management functions. This is justified by the fact that almost all subsequent mobility management proposals build, at least in principle, on the theoretical foundations established by Mobile IP4, known to-date as *IP Mobility management.*

Annex C.1 present a brief description of the Mobile IP(v4) protocol together with a signalling performance analysis.

Mobile IP and limitations

Since its inception, the primary concern for MIPv4 [101] has been simplistic connection transparency; this has been achieved with little to no effect on the existing IP addressing or routing infrastructure, sustaining thus, scalability. As shown in the previous section, such simplicity combined with path optimality, is afforded at the cost of design trade-offs, such as increased signalling load.

Subsequent MIPv4 revisions [102] improve on security issues pertaining to binding update authentication and implementation interoperability. While successful in achieving the above goals, honouring classic IP routing semantics, MIPv4 is designed to provide *elastic* IP handoff delay bounds. Its handoff management function assumes that the MN does not *change* its point of attachment more frequently than once per second. This is because during movement detection, encompassing agent (i.e. router) discovery, the MN cannot send an ICMP router solicitation message to the FA more often than once every second [103]. The limitation becomes more pronounced over increased error rates typical of a wireless link; in such cases loss of a solicitation message can stall the IP handoff of the MN for more than one second before it can re-issue a router solicitation.

The experimental work of Caceres and Iftode [104], has further exposed IP handoff *latencies* of 650 ms or more, once the MN associated with a new PoA and until its IP handoff completes. Their results suggest that elasticity of IP handoff delay bounds in MIPv4 comes not only as a result of the rate of IP handoffs, but also as a result of the mechanics of the process of IP handoff itself.

Yokota et al [105] confirm the observation by means of further experiments over 802.11b networks. They attest that the true total handoff delay varies between two and three seconds, with primary contributing factors, delay components of the L2 handoff process (AP probing), as well as IP movement detection. To alleviate the latency component incurred by the underlying link-layer, they propose a technology-specific optimisation to MIPv4 where the APs at the link layer forward received traffic for the MN during the period of the handoff. Their solution relies explicitly on special L2 devices that effect fast communication between previous and new associated APs to forward buffered traffic.

While such a solution is shown to perform satisfactorily, it remains technology-specific. It does not effect mere utility of link-layer triggers asserted in a number of other mobility management proposals; instead it relies on special hardware, network configuration of AP topologies and the underlying wireless link technology.

Mobile IP simplified under IPv6

Performance limitations as well as the patent encumbrance that Mobile IPv4 brought with its inception, rendered Mobile IP stagnant in progress or deployment. At the same time, the introduction of IPv6 as the next generation of Internet protocols whether wireline or wireless, set the mobility management function on an alignment course with the new IP protocol family.

Along with the advances of the IPv6 protocol family [106, 107, 108, 109, 110], Johnson and Perkins adapted the existing Mobile IP mechanisms through respective mobility extensions over IPv6 [111, 32], commonly referred to as Mobile IPv6. The underlying engineering of the core IPv6 protocols, allowed several simplifications to the mobility management function effected by the original Mobile IPv4 specification.

With respect to addressing, the IPv6 address has a default mode of stateless auto-configuration [110] or stateless/stateful configuration through DHCPv6 [112]. Both configuration types promote the model of *co-located* CoA address assignment to the MN. The abundance of identifiers in the IPv6 address space removes address assignment limitations for MNs; hence the routing can be effected end-to-end directly to the auto-configured CoA of the MN. From that perspective the Foreign Agent functionality[6] becomes redundant and was thus dropped from the Mobile IPv6 specification. Furthermore, the need for reverse tunnelling is eliminated by means of the *home address* option which is implemented by all IPv6 nodes; reverse tunnelling [113] has been required to eliminate the problem of ingress[7] filtering during upstream transmissions to the Correspondent nodes (CN). Eliminating reverse tunnelling effectively eradicates triangular upstream routing towards the CNs when ingress filtering is in effect at the visited link. A more elaborate set of improvements between Mobile IPv6 and its predecessor is presented in [32].

MIPv4 Agent Solicitations or Advertisements translate in IPv6 onto their Router solicitation or advertisement equivalents. In addition, these signals are extended to accommodate control flags to hint movement detection to the MN when transiting to a new IPv6 link. In addition, the transmission interval of Router Advertisement reduces significantly (between 50-1500ms) to aid expedite completion of the movement detection function.

2.2.5 Alternative Solutions to the Mobile IP(v6) doctrine

Beyond the mobility management model instilled by the Mobile IPv6 standard, a number of different approaches to IP mobility have also been proposed. These proposals view host mobility either through: (i) different layers of the network stack, (ii) multi-

[6]tunnel endpoint, relay

[7]host address as a source is topologically incorrect

homing (iii) name-based techniques. Each of these types of techniques has its own merits, introduces its own requirements and puts emphasis to different trade-offs. It is, thus, essential to overview different perspectives of managing IP mobility, to appreciate the relative or absolute importance of factors that may influence the design of the collective IP mobility management process.

In what follows, we present a brief overview of IP mobility management alternatives to MIPv6.

HIP

Host Identity Protocol (HIP) [114] transforms the security model of Mobile IP, and with it core MIP principles. Under HIP, identification is separated from location and routing through two identifiers: a permanent host identifier tag, represented through public keys as key hashes [115], and the normal IP address employed solely as a routing locator. The approach embeds the level of indirection between identifiers into a socket implementation of the network stack by introducing a host identity layer between the transport and network layers. In this manner all network connections remain bound to host identities, while the underlying dynamic bindings to assigned CoAs handle the routing of traffic between the MN and its peers. At the cost of modifying the network stack as well as the socket implementation in hosts, HIP eliminates authentication issues between communicating peers while it removes the need for tunnelling between the HA and MN, through IP-encapsulation or CN and MN through type II routing headers.

The underlying design of HIP simplifies significantly the issue of host mobility at the cost of complete reconsideration of the socket design and implementation. It is noted that IP handoff and address configuration are functions orthogonal to the mechanics of the MIPv6 protocol since they are defined within the IPv6 protocol core. Hence HIP eradicates the need for Mobile IPv6 simply by providing similar adjustments to critical IPv6 neighbour discovery functions. Although promising and a relatively fresh research direction in mobility management, HIP currently prohibits incremental deployment.

SCTP, TCP/mh

Stream Control Transmission Protocol (SCTP) [116] is a reliable transport protocol targeting acknowledged transfer of connectionless packet flows. TCP Multi-home (TCP/mh) [117] handles multiple local and remote address pairs in one TCP session, in the event that a single local/remote address pair goes down.

While the two approaches can handles multiple flows as part of a single transmission session, such as transmission to two different links, it may be seen that such approaches *cannot 'move'* a communication flow onto a new network. Such mechanisms can only *announce* new IP addresses for existing connections in the event the old connections cease to exist; they support no mechanism for migrating a network connection onto a new network if the previous connection has been broken, while the MN has reconnected under a new IP address.

SIP Mobility

SIP mobility [118] is an application-layer approach that provides support for real-time communications over mobile hosts. This is achieved by employing SIP [119] as control signalling between an MN and its peers. The scheme caters for simplicity by allowing the IP layer of all participating hosts to remain unaltered, while application-layer SIP messages are exchanged for host location, redirection or registration of the MN with its peers. The SIP mobility approach remains however severely limited in applicability since it cannot support TCP connections. What's more, SIP signalling introduces significant signalling delay during MN's IP handoff or MN registration with its peers; however, it also prevents the formation of triangular routing experienced under standard Mobile IP.

Naming

The notion of *naming* in the Internet, has referred, typically, to either an IP address or a domain name service (DNS) name. DNS is used, typically, to create mappings between domain names and IP addresses [120]. The idea behind DNS is that of associative identification; humans recall names better than addresses. As such they can identify an entity faster by a name than by a (large or unstructured) number. In addition, with DNS it is possible to maintain the same domain name while changing the nodes underlying IP address (either due to host failure, host service replication or change in the network interface. The latter, however, may also be manifested as physical change in the network (IP) subnet location of the host, while still maintaining the original name.

The *Real-time DNS* (RT-DNS) proposal [121, 122] aims to extend DNS with real-time update capabilities by actively flushing DNS caches. Such approach, requires that all DNS servers in the hierarchy support such functionality with updates propagating all the way to the DNS root(s).

The RT-DNS algorithm tends to concentrate the update the root servers, since all MNs must initiate ripple DNS update waves that terminate at the root of the hierarchy [123, 124, 125]. This mechanism cannot provide by itself transparent communication between the MN and its peers; it deals only with fast mapping of the MN new IP address onto its domain name. Thus, an additional connection mechanism such as SCTP [116] or TCP/mh [117], is required to announce the new IP addresses to (socket) connection or network state maintained at MN's peers.

According to [122] delays in the order of five seconds can be imposed in updating the authoritative DNS chain per IP handoff. Such performance is clearly limiting for the performance of interactive applications over IPv6 mobile network infrastructures.

With the exception of HIP, most of these approaches are significantly limiting and hence expected to meet equally limited acceptance. HIP itself is a rather new[8] research direction emanating from peer to peer networks [126] with a number of open research issues [127, 128], that do not deal with the core routing function for mobility management purposes. To this end, further investigations with HIP are beyond the scope of this study.

A set of alternative naming approaches in regards to a mobile (albeit ad-hoc) node are further provided in Annex C.2. This set of schemes demonstrate a different angle of attack on the issue of naming for the combined approaches of both host identity and routing for the purposes of mobility.

The following section presents a challenges arising with the introduction of Mobile IPv6 as the de facto IP mobility management standard.

2.3 Macro-mobility: emerging issues and challenges

In the evolutionary reality of science, solutions to a problem point, as a result, to the *evolution* of the original problem, in contrast to the apparent expectations for a *complete* solution. Irrespective of how disciplined the solution to a problem is, when attached to refined visions of applicability, it produces new input either in the form of solutions or requirements. The latter augments the information space about the problem generating fresh *aspects* of the problem. Hence, in absolute terms, one can never solve the original problem, but evolve it!

In a similar fashion, Mobile IP - whether v4 or v6 - while addressing connection transparency, 'fails' to preserve connection *delay* transparency during MN's IP handoff, compared to stationary wireless *or* wire-line IP communications. In this sense, the IP mobility management function is said to lack (delay) *seamlessness* in its control deliberations.

[8] HIP has come into existence only during the last 1-2 years

To bring semantics into context, *seamlessness* may be viewed as the ability of the network to support transparent[9] IP mobility between consecutive points of attachment. Such transparency is manifested as **low** delay and associated packet loss.

From the above it can be seen that, the reason justifying such 'failure' is in fact *two-fold*; support for seamless mobility was *not* the primary design consideration in Mobile IPv6 proposal [130]. In addition, Internet has been by-design a *best effort* network; the humble beginnings of Internet did not consider mobility or its seamless descendant, at the outset, despite the first research endeavours in the past over Packet Radio [131]. Thus, the lack of seamlessness in the collective IP mobility management function does not come as a surprise to the network engineer.

The modern mobile internaut however, is not concerned with that. Cellular networks have set the communication paradigm; the average mobile user becomes now more concerned that his mobile computing device *sustains continuous connectivity* to some wireless IP network infrastructure. Such user requires freedom from reconfiguration or disruption in the engaging activity, encompassing high-level IP application services. Seamlessness and transparency become essential in the user's mobile IP network communications.

The following sections present issues sensitive to the application of IP mobile access practices over wireless network infrastructures. Identification of an exhaustive set of such challenges is beyond the scope of this thesis. To this end, we focus on issues and challenges relevant to subsequent investigations in this study; these set the foundations of our investigation and re-emerge as they are being addressed during experimental or stochastic simulation measurements in subsequent chapters of this thesis.

2.3.1 Interactive IP Multimedia and hard delay bounds

Interactive IP applications, involve a two-way (or multi-way) transmission of UDP packet flows between participating hosts whether fixed or mobile. Such IP applications impose hard real-time packet delivery constraints, particularly[10] on end-to-end delay.

The maximum acceptable latency depends on the type of application. Karlsson [35], and Kurita et al. [132], show that one-way delays of 100-150 ms are acceptable when assessing the perceived quality of a single IP telephony (VoIP) flow. For audio, a latency of 100ms is identified as the hard (one-way) delay bound beyond which a

---

[9] A more elaborate definition has also been provided in our work in [129]

[10] other constraint is bandwidth

human can recognise a pause before it initiates a talk-spurt response [41]. Above[11] 150 ms delay renders the quality of interactive audio/video traffic significantly degraded and in many cases unacceptable [24].

For the purposes of IP mobility management, delays incurred as a result of the type of the audio/video encoding are very small and thus do not influence VoIP performance during a handoff. Despite that different audio/video codecs have somewhat different delay requirements, and as such impose different requirements in terms of one-way delay [133], our investigations assume a simplified uniform delay bound of 200ms.

Furthermore, for interactive IP applications, the measure of associated packet loss may be influenced by the type of audio/video encoding adopted. This is because each encoding generates constant bit rate (CBR) flows with a different packetisation rate. For instance, popular speech encodings such as A-law, μ-law, or G.711, used in IP telephony, generate packetized audio samples of 20, 40, or 80ms respectively. Thus, within a fixed delay period a low packetization rate (e.g. 80ms) incurs small packet loss, with each packet containing a significant part of the voice utterance; on the contrary a high packetization rate incurs a higher packet loss, with each packet payload containing a smaller part of the talkspurt.

Steinmetz has shown that any form of delay variation (*jitter*) affects the intelligibility of interactive audio/video to the extent that the interactive communication style must change [25]. This occurs for jitter values in excess of 100ms. For combined interactive audio-visual applications (videoconferencing) the effects of increased jitter are made apparent to the user as loss of stream (lip) synchronisation [134, 135]. Jitter is an additional component of latency owed typically on the build-up of routers' forwarding queues and detracts from the timely delivery of real-time IP traffic to end hosts.

For other interactive IP applications such as network games, acceptable latencies differentiate with respect to eye movement and motor reactions. Vision acts as an oculomotor integrator of input stimuli [136], whereas hearing acts as a differentiator [137]. The process of stimuli integration (vision) yields motor reactions with smaller latencies than that of stimuli differentiation (audio); this is because signal composition (integration) for the visual modality can be achieved from less complete input stimuli than does the auditory modality through the function of signal decomposition (differentiation).

For vision, Cheshire indicates the minimum latency component to be the raster scanning of single on-screen frame, which is 33ms [41]. With respect to motor reactions,

[11] more accurately at 200ms

Maki et al. identify that user tolerable latencies depend on the stability of the responded stimuli of limb expected reaction times [138]. In particular, while the response of the limb could be as fast as 50 ms, the stability of the task achieved was very poor, i.e. the response did not contribute to a successful action. However when the response of the limb delays as much as 135 ms the task stability reaches about 80%; that is, 80% of the tasks effected by the limb response, are achieving the target objective. The stability of success of the task reaches 100% for limb reflex response of 270ms. While the experiments were carried out on lower limbs (legs) while most network games are played by hands, it can be argued that there exist more correlation of the limb motor responses between legs and hands rather than voice talkspurt latencies tolerated according to hand reflex responses.

The above indicate that, delays of 270ms may be considered acceptable for network gaming applications. Classic human factors research [139] argues that 200ms is a more acceptable time for latency tolerance on the user, while MacKenzie et al. [140] and Park et al. [141] establish that Virtual Reality interactions become difficult after 225ms of latency.

With this in mind, the growing volume of UDP real-time traffic (audio/video streaming, IP telephony, conferencing, games or other) brings about a cause for attention towards delay-sensitive considerations in IP mobility management system design. To this end, the definition of IP mobility seamlessness is aligned to denote the measure of end-to-end network delay within acceptable bounds for interactive IP applications.

It should be noted that the notion of seamlessness for the purposes of mobility must be distinguished from the notion of resource reservation with respect to the Quality of Service on the transported real-time traffic. The former refers primarily to the reduction or elimination of any latencies that can induce disruptions or packet loss on the transported IP packet flow, to or from the MN, irrespective of its mobility pattern. This does not provide any qualitative guarantees with regard to congestion in the network; it simply attempts to remove any mobility-related factors that can augment the end-to-end delay experienced by an IP flow as a result of IP movement.

2.3.2 Increasing Rate of IP handoffs

Cell size in wireless systems decreases with an increasing demand for link capacity, power control and accuracy of location information [142]. From a technology perspective, reductions in cell size are further confirmed by the popularity of 802.11 WLAN networks [143] and Personal Area Networks [144] realised through Bluetooth and IEEE802.15

[145]. These types of wireless networks offer (or expected to) picocell coverage on a per link basis with ranges around 10 meters. In such networks, IP services are available on smaller coverage radii of wireless last-hop links in comparison to cellular networks.

For mobile node trajectories cutting through such coverage areas the net effect is shorter residence times. Proof for such observation can be readily derived by looking at the relationship of cell residence time and a random straight line trajectory cutting through a circular coverage area, as illustrated in figure 2.2.

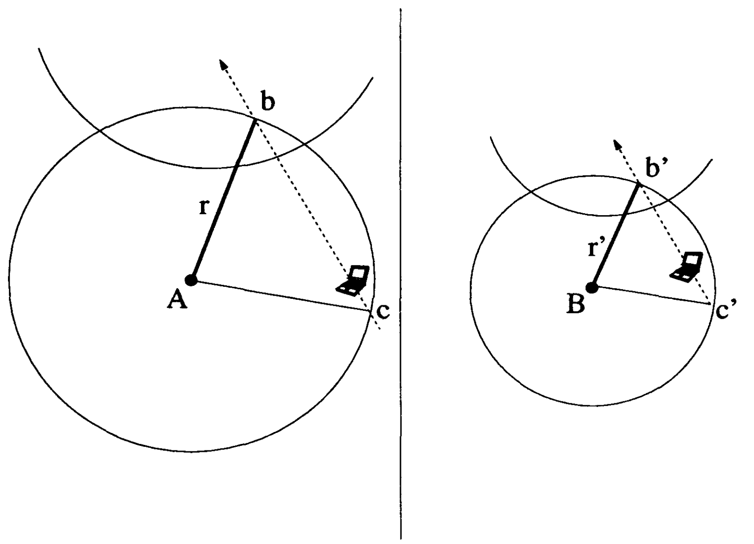


Figure 2.2: Mobile Node IP handoff interval affected by cell radius

The chord bc representing a simplified trajectory of the MN through a cell may be represented as:

$$cb = 2R\sin(\frac{\theta}{2}) \tag{2.1}$$

with an upper bound of $2R$ when the chord becomes the circle's diameter. Thus, the simple relation between intra-cell travelled distance and velocity becomes

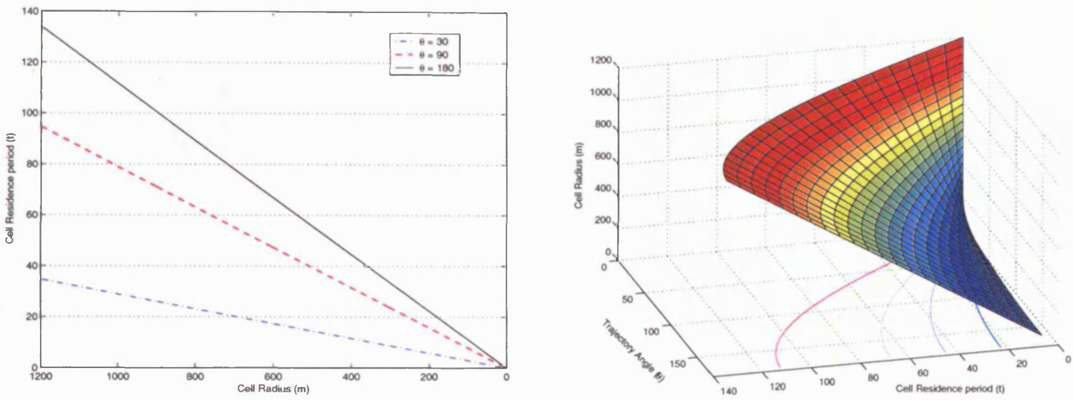
$$t_c = \frac{cb}{v} = \frac{2R\sin(\frac{\theta}{2})}{v} \tag{2.2}$$

where t_c denotes the cell residence period or equivalently MN's handoff interval; the later is inversely proportional to MN's handoff rate, i.e.

$$h_r = \frac{1}{t_c} = \frac{v}{bc} = \frac{v}{2R\sin(\frac{\theta}{2})} \tag{2.3}$$

IP handoff interval (t_c) and cell radius R has a linear relationship as shown in figure

2.3(a); for an increasing coverage radius, MN's handoff interval increases proportionally, assuming identical trajectory angles. This is not the case for handoff interval as a function of the trajectory angle with respect to cell's centre. Figure 2.3(b) shows graphically the effect of intra-cell trajectory angle on IP handoff interval as a function of cell radii; we observed a quadratic growth of the handoff interval period for increasing trajectory angle with respect to the cell's centre.



(a) radius vs. handoff interval

(b) radius vs. handoff interval versus traj. angle

Figure 2.3: IP Handoff interval as a function of cell radii and trajectory angle

The above imply that for a decreasing cell radius, the expected handoff rate will increase as the MN is travelling at constant speed between consecutive cells realising IP access networks. Quadratic growth or decay of handoff intervals is expected for fluctuations in trajectory angles.

In the light of increased mobility as a result of smaller cell size, Pollini et al. projects subsequent increases in signalling traffic rates, confirming also the tendency for higher mobile device density and handoff rate per cell[146].

Beyond Horizontal IP Handoffs

The initial state of maturity in cellular networks accompanied with growing acceptance of 802.11 WLANs have pushed the evolution of future wireless systems towards *hybrid* wireless network configurations [147]. A hybrid wireless network embraces a combination of macro-, micro- or pico-cellular infrastructures deployed within the bounds of a geographical area. In this manner, individual strengths of each wireless technology or domain infrastructure can be exploited [148]. Thanks to the tether-less character of the air interface, each individual[12] wireless technology - current or emerging - comprises a

[12] or Wireless Internet Service Provider (WISP)

network overlay within such hybrid network [149].

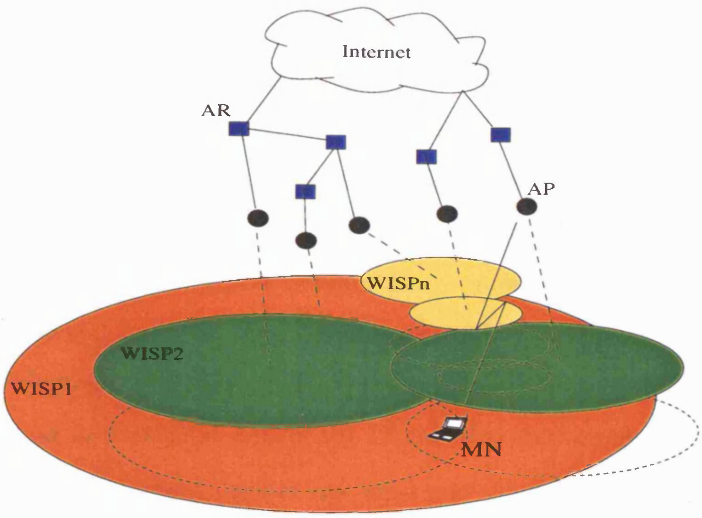


Figure 2.4: MN facing numerous possibilities for vertical handoff, irrespective of mobility

However, the availability of multiple network overlays within a given geographical location provides the MN with numerous (instantaneous) possibilities for IP handoff, irrespective of its mobility (i.e. velocity). Figure 2.4 illustrates a hybrid network model, whereby the MN is presented with the possibility for multiple IP handoffs, not as a result of movement but of *availability*. In such deployment scenarios, the MN is presented with the potential for a *vertical handoff*.

There has been an abundance of handoff schemes in the research literature [150, 151, 152]. However, in their vast majority these proposals have focused on horizontal handoffs across homogeneous[13] points of attachment. More important, the underlying algorithm makes use of properties met at the link or physical layer of the supported wireless medium. Typical examples of such properties or functions have been the measure of signal-to-interference ratio (SIR), coding schemes, transmission frequencies, power control, or the site topology of the wireless access network, shaped as a result of the MN's mobility.

On the contrary, the possibility of a vertical IP handoff does not necessarily indicate movement; instead it indicates availability of multi-technology networks or multiple competing WISP with different provisioning *capabilities*. For instance, cellular network WISP overlays can provision user bandwidth that ranges from 50 Kbps (GPRS), 100 Kbps (EDGE) and 100-384 Kbps (UMTS), with smaller bandwidth, but comfortably

[13] The same technology across all APs

larger coverage; on the contrary 802.11a/b/g WLAN networks can provide significantly larger aggregate bandwidth ranging between 1 and 54 Mbps, but significantly smaller coverage.

The above, in effect, can translate to a vertical IP handoff between wireless IP domains, not as a result of link layer characteristics, but as a result of service diversification and ultimately, user *choice*. From this perspective, the definition of handoff rate realised as a function of cell radius, velocity and angle of trajectory has little or no meaning. The handoff rate becomes now a function of potentially *measurable* user choice. Thus, through vertical handoffs, the handoff interval is guaranteed to vary far more dynamically. This presents an important call for consideration of handoff management at the IP layer and in particular within the IP mobility management function.

The mobility management mechanism must now consider the fresh possibility of a significantly higher handoff rate; in parallel, encompass the function of *facilitation* and provisioning of advanced IP handoff criteria or metrics to the mobile node concerning the network - existing or future - and its services.

2.3.3 External latency factors

Despite the improvements effected in Mobile IPv6, the protocol standard has significant limitations towards support for seamless delivery of real-time traffic, namely minimum delay and associated packet loss.

With respect to control signalling, an important set of contributing factors influencing MIPv6 performance is delay incurred beyond the control of either Mobile IPv6 or the underlying core IPv6 protocols. Such class of delay factors is collectively identified as *latency externalities* with respect to IPv6 mobility. The following sections present briefly important measures of latency externalities and their effects on IP mobility management.

Packet delay characteristics

Bovy et al provide a comprehensive account of of end-to-end delays over a large number of paths under the auspices of RIPE Test Traffic Measurements (TTM) project [153]. The study acquired readings from about 40 measurement nodes around Europe with a number of nodes on US transatlantic link and New Zealand, measuring one-way delays over 963 paths. The large number of paths offers a significance level of confidence on end-to-end packet delay characteristics.

The study reports that the vast majority of end-to-end packet delays is well approximated by a shifted gamma distribution. The reported mean delays are 14.5ms and 110ms for the *local* (i.e. intra-domain) and the *transatlantic* (i.e. inter-domain) paths, respectively. The study emphasises that the delays reported show very sharp peaked distributions, with most delay values clustered within about 10% of both the mean and minimum values. The former implies that excessive[14] delays are in reality quite rare. Another important observation is that the transatlantic path actually exhibits substantially smaller standard deviation (3.085) than that of the local path (6.216). This is suggested to account for smoothing effects produced by a relatively large number of router hops over the international route.

The above empirical end-to-end delay study has a number of implications for the IP mobility management function and associated handoff performance; in particular, it appears that the above means employed through a shifted gamma distribution can be used to model *realistically* intra and inter-domain end-to-end path delays between the MN, its HA and its communicating CN peers.

Congestion and End-to-end delay The measure of end-to-end delay over IP mobility has been reported as another contributing factor by Mukkamalla and Raman [154]. Their study investigating the amount of latency due to congestion, report a typical one-way delay of 70ms for no congestion, with heavily tailed peaks of 1600ms during congestion; running day-time latency averages were reported with a mean value around 540ms and median value of 108ms [155]. This implies a left-skewed heavily tailed one-way delay distribution, which confirms the TTM results.

An important observation is that for lost registration replies one-way delay peaks at 1950ms. Their study also points out that peak latencies account also for the limitation of the HA to tunnel traffic up to 2500 hosts; latency begins to soar above this figure. Considering the capabilities of cellular networks for handoff management at the rate of 3000 h/sec, the figure become representative of the capacity for a single home agent.

It is important to note that host capacity at the HA is also a function of the packet size; the reported HA capacity below which the HA can manage traffic tunnelling with acceptable end-to-end delay corresponds to maximum packet size of 250 bytes. For smaller packet sizes (160 bytes), the MN host capacity of the HA increases up to 20000; it implies that a 40% decrease of the maximum packet size handled by the HA, incurs an increase in tunneling capacity by about 0.8 orders of magnitude.

[14]more than 2 standard deviations above the mean value.

Round Trip Time

Round Trip Time (RTT) is a composite external delay characteristic relevant to IP mobility management; it arises from asymmetric end-to-end delays experienced in the forward and reverse path of the end-host. Round trip time delay is independent of the transport protocol; its one-way delay component is primarily attributed to queue build-up due to congestion on intermediate links of a routing path towards a destination host.

Jiang and Dovrolis's study of RTT distribution, for TCP purposes, attests that more than 90-95% of probed connections have an RTT that is less than 500ms. In the case of regional links, more than 75-90% of connection have an RTT less than 200ms [156]. In terms of the RTT bounds, they show that about 35% of TCP connections in all traces exhibit an RTT of 50ms or less, while the remaining connections experience an RTT of up to 200 ms. However, these connections experience a significantly lower[15] number of hops, as their end-points are located within the local geographical area of the monitored link.

The authors indicate that backbone links have a wider RTT distribution than that of access links; core links carry traffic between multiple geographically dispersed populations of users in comparison to access links carrying traffic from the regional user population.

Correlation of RTT and number of hops In general, the mean value as well as distribution of RTT, at a link, depends on the geographical location of the connection's end points [156]; for instance when two end points are found to be within the same administrative domain (typically over a single geographical region), RTT has a smaller mean value and less heavily tailed distribution; the opposite is the case for two end points residing over different network domains connected through the Internet core. Hence different sets of links may have significantly different mean RTT values and distributions as they depend on localised effects of congestion.

Crovella and Carter measure RTT distribution as a function of the number of hops traversed across a link from 5262 paths to respective WWW servers [157]. They report that the number of hops is normally distributed with a mean of $\bar{x}$=16.6 hops and standard deviation $\sigma = 4$. To assess differentiation between regional (intra-domain) and backbone (inter-domain) paths, they classify paths according to the size of their

[15] The lower RTT bound at a monitored link cannot be less than the round-trip propagation delay of that link.

administrative domain. They find that for intra-domain paths within the particular domain, the number of hops remains normally distributed with a mean of $\bar{x}$=8 hops and a standard deviation of $\sigma = 3$.

The authors emphasise that distribution of associated RTTs is markedly different from the aforementioned distribution of hops for each monitored path; in contrast to the normal distribution of hop count, the RTT distribution is strongly left-skewed reporting a median of 125ms ($\bar{x}$=241ms), which is characteristic of the exponential or gamma distribution. This is confirmed also by RTT standard deviation (about 2$\bar{x}$), compared to the hop count std. dev. (about 0.25$\bar{x}$). Such differences indicate that RTT and hop count in a communication path are not strongly correlated; only a 10% of the total variation can be explained by a linear relationship between hops and round trip time [157].

Higher RTT-hop correlation is reported by Gwertzman and Seltzer [158]; however results appears to be derived under limiting assumptions, while the reported correlation itself is not significantly strong.

The above results are also confirmed by Acharya and Saltz [159] while they deviate slightly, suggesting that the mode is a better value characteristic of the RTT distribution. The authors report further that RTT distributions change slowly. In particular, persistent changes in RTT occur slowly, while sharp variations are eliminated quickly.

The RTT distribution results of Crovella et al, are validated by work by Guyton and Schwartz who report a mean of $\bar{x}$=17 hops and variance of $\sigma = 4.3$ [160].

RTT variability and implications for IP mobility The reported RTT results indicate first a high variability of delay between intra-domain and inter-domain paths. In a ubiquitous wireless Internet this translates to non-deterministic delays on end-to-end signalling; this is of particular importance to design and subsequent performance of the IP mobility management (IP MM) function.

In particular, variable RTT delays may impose adverse performance effects on IP MM signalling that manages the critical period of MN's IP handoff. This is particularly the case when the continuity of packet communications is critically dependent on the completion of the IP handoff process. Such is the case, for instance, with Mobile IPv6 where binding update signalling sent towards MN's HA and communicating CNs during an IP handoff, conditions the continuity of packet flow towards the MN.

Subsequent chapters of this thesis identify accurately performance limitations relevant to external delay factors influencing MN's IP handoff; we subsequently explore how to overcome such signalling dependencies for critical IP MM performance under interactive application service requirements.

Wireless Medium Access Control

Wireless multi-access channel systems are based on the principle that hosts acquire the medium when they want to effect a frame transmission. To this end a wireless link technology implements two sub-layers: (i) the physical layer and (ii) the link layer. While the physical layer allows the modulated transmission of an encoded bit-stream over the air interface, the link layer effects an ordering commonly referred to as medium access control (MAC).

Most media access control protocols for multi-access channel systems enforce a coordination function - whether distributed or centralised - whereby the wireless station (STA) *contends* for medium access amongst other STAs, including the Access Point (AP) itself. In an infrastructure wireless IP network, expected to accommodate large number[16] of wireless hosts, the wireless link at a PoA, is expected to experience increased contention for medium access.

At the same time, for a wireless multi-access system implemented on a single channel, frame collisions from attempted transmissions can be prevented not by means of *detection* [161], but by means of *avoidance* [162]. This is primarily due to the *cost* of full duplex wireless interface implementation and the *overheads* imposed on the available bandwidth supported by the wireless medium, over a single (multiplexed) channel. To this end, a number of MAC protocols have been proposed [163, 164, 165, 166]; amongst them perhaps the most popular is Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) [167]. Under this protocol, the node verifies momentarily the absence of transmission activity before acquiring the medium for its own transmission. In the event that the medium is busy, the wireless station is forced on a delayed reattempt to sense and capture the medium. The delay enforced typically follows a binary exponential increase up to an upper delay limit[17].

It may be seen that as a result of a densely populated wireless link with increased offered traffic load, communicating MNs are expected to experience an increased transmission delay as a result of contention for medium access. As a result the end-to-end delay of a wireless host may increase *not* as a result of *congestion* on intermediate hops in the path towards the CN destination, but as a result of *contention* for medium access

[16] in the order of hundreds or thousands.
[17] to avoid starving the host from transmitting by waiting indefinitely.

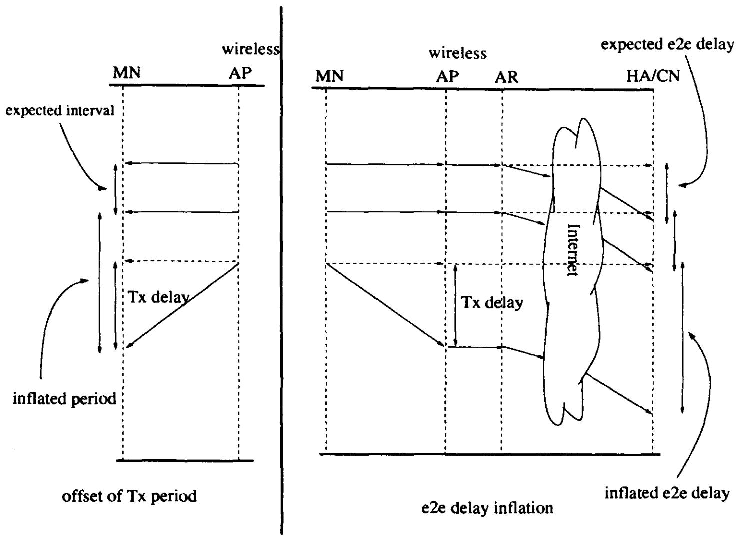


Figure 2.5: Influence of wireless link delay as a result of medium access contention in densely populated Points of attachment

on the wireless link[18].

While an increase of the end-to-end delay as a result of contention may be acceptable for packet transmissions, it is expected to have a significant impact on control signalling effected by management protocols, particularly for IP mobility. The significance lies in the fact that contention in wireless links effectively *offsets* the transmission of signalling by some delay period as shown in figure 2.5.

The above has a cascading effect on two aspects of control signalling: (i) periodicity of the control signal (ii) end-to-end delay or RTT inflation in the transmission and consequently receipt of the control signal. These effects can have a significant impact on the time of IP handoff completion, particularly for IP application services with stringent delay requirements, such as interactive IP applications. Chapter 3 exposes the effect of latency as a result of network-level signalling delays and link-layer contention, during interactive communications over 802.11b networks.

Currently, existing IP mobility management standards such as Mobile IPv6, *do not* consider wireless link delay as a result of contention in their systems design. The primary reason is that most proposals strive for layer independence in their protocol design; this allows independence from assumptions about the underlying wireless technology.

However, from the above it is clear that the IP mobility management function *must* encompass such performance consideration in its performance requirements, if it is to

[18] In fact such limitation is expected to emerge not only on last-hop wireless links but also on intermediate hops that in future may attempt to furnish point to multi-point bridges in the L2 fabric of routers. In todays networks this is not the case as most wireless links are point-to-point

meet stringent delay guarantees required for interactive communications. We argue that such considerations can be encompassed without particular dependence or assumptions about the underlying wireless link layer.

2.3.4 Latency and State Establishment

IP address allocation

Snoeren et al identify in [168] that perhaps the most significant problem incurred in mobility is the latency induced during IP (re)connection. Finney and Scott report in [169] that the allocation of an IPv6 address requires a minimum of 1600ms assuming no DAD [110] hits.

With respect to stateful address configuration, Vatn [170] has shown the effect of latency incurred from the IPv4 address allocation on interactive multimedia; DHCP-specific address configuration require 10-15sec without ARP checking [171], 5 seconds with ARP checks suppressed or 150ms with alternatives such as DRCP [172]; these are figures derived from real-world implementation experience for mobile multimedia [173].

For IPv6, the stateless address auto-configuration mechanism in IPv6 mandates the verification of uniqueness for the address configured by the MN. Reason for that is the intentional or unintentional configuration of duplicate IPv6 addresses on the same link. While the probability of unintentional configuration of an IPv6 address is statistically extremely remote[19], the same cannot be assumed for cases of malicious configuration of a duplicate IPv6 address, whether by fixed or mobile hosts. The latter arises in for various types of security attacks stemming from IP address spoofing [174]. As a result it can be seen that the process of duplicate address detection (DAD) cannot be ignored on the grounds of statistical improbability. To this end, address establishment may need to be generated cryptographically by adapting the security mechanisms of neighbour discovery [175]. Despite such adaptation, however, the duplicate address detection function remains an integral part of the address auto configuration process.

The above implies that the allocation of IPv6 addressing state on itself generates by default enough latency to place any active IPv6 flows well beyond the boundaries of acceptable guarantees for real-time traffic delivery. This is a significant call for consideration in the design of the IP mobility management function.

[19] $1/2^{64}$ for an interface identifier of 64 bits

AAA or QoS allocation

Operational deployment of IP mobility management over wireless access networks at a commercial grade, requires that service providers can provide certain quality of service (QoS) guarantees; in addition, commercial viability dictates the enforcement of basic authentication, authorisation and accounting (AAA) functions, for access control and revenue accrual.

From the perspective of the network layer the MN must attach to a new IP point of attachment (PoA) and configure a new IPv6 address. Admission, however, of the MN at the new PoA is subject to authorisation and authentication for billing purposes. Currently there is no *standard* AAA mechanism in support of the IPv6 mobility management function.

AAA [176] and QoS [177] extensions to Mobile IPv6, currently under development, employ different sets of signalling interactions incurring an additional 0.5-1 RTT in any existing signalling delay. Depending whether binding registration state is authenticated by the farthest[20] CN, reactive exchange of AAA state brings the gross latency total between 2-3 RTTs; this is before the MN is granted packet transmission privileges and can receive its first data packet at the new IP PoA from that CN.

Furthermore, to assure a certain level of Quality of Service during Integrated QoS service provisioning, the entire end-to-end path needs to be re-established at the new IP point of attachment [178]; that is, each and every router on the path between the MN and its CN peer, must be configured to effect the required level of service quality[21]. This adds one extra RTT to the total latency before the active IP flow of the MN can resume communications with its peers.

With reference to AAA and QoS, acquisition of such multi-context IP connectivity state, is currently effected sequentially in reaction to detection of movement of the MN between consecutive points of attachment. Under the existing IP mobility management models, the MN must first acquire it IP addressing state **before** AAA functions can take place, since authentication and accounting must be based on MN's IPv6 address.

In addition, there appears to be an implicit ordering on the set of state contexts established; revenue accrual warrants AAA state establishment **before** the MN is granted access permissions and subsequently any kind of QoS for communications with it peers.

---

[20] in the case of multi-CN communications with the MN

[21] alternatively crossover QoS path repair can be effected; however this requires that first that the crossover point between the old path and the new path is established before the crossover QoS path leg can be repaired

Cascading reactive establishment of different context-state, ultimately incurs a multi-RTT transmission delay for the MN as it transits from its current to a new IP point of attachment. Under existing mobility management standards, the magnitude of such transmission delay is responsible for the lack of seamlessness in an IP handoff between consecutive points of attachment.

2.4 Reconsideration of Macro-mobility

Along with the developments of micro-mobility management protocols presented in Annex C.3, a number of *macro-mobility* extensions to Mobile IPv6 has been emerging recently, in support of seamless IP mobility management for interactive multimedia services.

These MIPv6 extensions aim to address issues and challenges pertaining to delay seamlessness, as described in Section 2.3. At the time of writing, these mechanisms are the object of on-going research effort in the respective engineering[22] and research[23] working groups of the IETF and IRTF.

In parallel with the requirement for delay transparency, future visions of multi-network overlay infrastructures [179] or operational performance requirements present fresh challenges pertaining to:

- fast handoff management: addressing most or all delay-prone functions of the IP handoff process in control of the IP layer.
- security of mobility management signalling: to address growing concerns about denial of service (DoS) attacks due to unauthenticated signalling.
- header compression: to address bandwidth conservation issues emerging from small packet size flows, such as the ones of IP telephony.
- context transfer: to address efficient forms of relocation or establishment state at new points of attachment, in aid of expedited multi-context IP handoffs.
- movement detection: to address efficient forms of detecting the movement of the MN between points of IP attachment in aid of expedited *movement-based* IP handoffs.
- protocol scalability: allow for distributed mobility architecture that prevents performance bottlenecks, while minimising signalling overheads

[22]Such groups encompass the Signalling and Handoff Optimisation (MIPSHOP) WGs of the IETF
[23]These encompass the: IP Mobility Optimisations (MOBOPTS) WG of the IRTF

- QoS capabilities: cater for QoS provisions as part of the overall mobility management function.

- incremental deployment: allow for gradual deployment of the supported mobility management function in the underlying network infrastructure as a result of expansion.

It can be seen that the core of most issues emerging for current or future applications of the mobility management at the IP layer, emanate from the requirements of preserving seamlessness in two main research fronts:

1. *enhancing* components of the *IPv6 handoff* process. Proposals from this research perspective fosters brute-force techniques that alter or enhance existing functions of the IPv6 handoff process to minimise or eliminate delay from the IP handoff process.

2. promoting *forwarding transparency* during MN's handoff. In this area, research investigation focuses on ensuring that traffic destined to the MN is forwarded efficiently at the MN new point of attachment to prevent disruption in flow communications in view of MN's IP handoff.

In the following sections we present briefly current research efforts that attempt to address in different forms, the above performance challenges. We note that IP mobility management is currently a very active research area; current on-going research efforts have been addressing independently the above mentioned seamlessness issues. As a result, proposals presented in the following section are subject to constant change and evolution and thus, expected to change.

2.4.1 Expediting IP handoffs

A number of macro-mobility proposals belong to the family of *fast handoff* protocols. Their purpose is to minimise handoff delays as well as the disruption in packet flow experienced by the MN while moving between successive points of IP attachment. The following sections present current research efforts in support of delay seamlessness for the IP macro-mobility management function.

Tunnel-based (unicast) forwarding: Fast Mobile IPv6

Fast Mobile IPv6 (FMIPv6) has been a recent research effort in the IETF aiming to minimise the handoff latency of MIPv6 [33].

To enhance the handoff management and flow forwarding function of MIPv6, the FMIPv6 proposal requires link-layer information from the underlying wireless technology at hand. This is done for two reasons: (i) aid the configuration of the care-of address of the MN on the new AR before MN upcoming IPv6 handoff; (ii) establish a unicast tunnel between the previous and new point of IP attachment to effect flow forwarding during the MN's IPv6 handoff.

From that perspective, it appears that an FMIPv6 solution is expected to be link-layer specific, since link-layer information are guaranteed to be heterogeneous or experience different performance across different technologies.

FMIPv6 distinguishes between two types of handoff: (i) reactive (ii) anticipated or predictive. Both cases of handoff types are tunnel-based, that is, FMIPv6 enforces tunnel-based forwarding between the previous AR and the new AR while MN's handoff is in progress.

A *reactive* FMIPv6 handoff refers to the case where the MN *breaks* its association with the current point of attachment *before* it can *make* an association with the new point of attachment (PoA). An *anticipated* FMIPv6 handoff refers to the case where the MN can *make* an association with a new point of attachment *before* it *breaks* its existing association with the current PoA. In both cases the MN may configure an IPv6 address of ahead of its handoff, whether reactive or anticipated, by means of information conveyed through FMIPv6 RtSolPr/PrRtAdv messaging.

Anticipation of the new PoA is effected by means of link-layer (L2) triggers. An L2 trigger is information pertaining to the underlying link-layer protocol, utilised to initiate a network-layer handoff before the end of the upcoming link-layer handoff from the current PoA. It contains information about the link-layer connection as well as identity of participating APs. An anticipated FMIPv6 handoff is shown in figure 2.6

Recent experimental work Montavont and Noel [180] reports that handoff anticipation under FMIPv6 can be erroneous or imprecise resulting increased handoff delays. The authors establish as a significant cause the fact that the MN *cannot* use its newly formed Care-of address until it has been acknowledged by the new AR before or after its IP handoff transition.

They report a pure IPv6 handoff latency of 151ms for the reactive type of MN IPv6 Handoff and around 186 ms in the case of the L2-anticipated IPv6 Handoff; these values are independent of the underlying L2-handoff delay. The anticipated type of

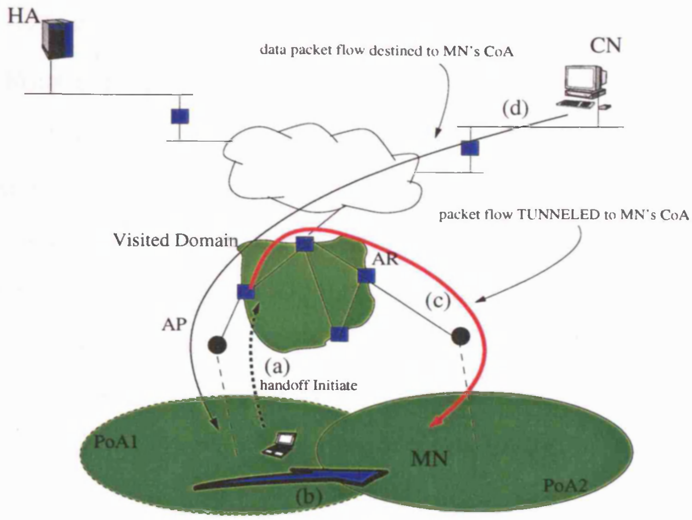


Figure 2.6: FMIPv6 tunnel-forwarding operation in response to L2-trigger-based handoff initiation

handoff appears to incur more than the case of a reactive handoff; this is the result of early establishment of the tunnel with the new point of attachment with respect to the actual initiation of the de-association of the MN from its current point of attachment.

The authors present also FMIPv6 handoff latencies in the face of lower signalling rate and for 3-6 wireless hosts communicating on the same wireless link with wire-line peers; lower signalling rates arise typically at the boundaries of a coverage area or as a result of adverse propagation effects. For the reactive case of IPv6 handoff the authors report a handoff delay of 1723ms, while in the average case of the anticipated IPv6 handoff is around 7235ms; the case of anticipated handoff appears to be almost four times greater than the case of a reactive FMIPv6 handoff. This appears as a result of significant dependency of FMIPv6 signalling on resolution of medium access control (MAC)-based contention before signals can be propagated between previous and new ARs. It is noted that transmission delay for the communicated packets between MN and CN in their experiments is *not* significant to introduce *any* packet losses in post-handoff communications.

While FMIPv6 is currently the subject of on-going work in the IETF the above results suggest that the existing signalling of FMIPv6 does not account for the fact of MAC contention, in the face of two or more MNs associated with the a point of attachment. Clearly the above mentioned delays cannot act in support of interactive application service such as IP telephony or videoconferencing over wireless networks, in particular 802.11 WLANs. A comparative performance analysis of the FMIPv6 propsal is presented in Chapter 6.

Multi-tunnel Forwarding

Lee and Jung et al [181] proposes an adaptive resource allocation mechanism that aids handoffs in terms of speed and reliability. The scheme adopts SNR thresholding beyond which a list of Base Stations already formed by the MN is communicated to the master base station (BS). Below a critical SNR threshold the MN initiates a handoff request to the target BSs with its own address as well as bandwidth reservation information. Upon registration of target BSs with the BS master, the latter forwards packets to these BSs. The scheme is assumed to forward the packets to the target BSs over multiple unicast tunnels, since it does not cater for multicast mechanisms, as the latter require in-advance group membership setup which is not considered at all as part of the protocol. The main deficiency of this proposal is that the MN is expected to contact the target BSs first for the handoff setup phase. In the face of propagation effects the MN is required to contact each target BS individually. Also there is no specific consideration of the optimality of the SNR thresholds for handoff initiation and BS group updating with the master BS. This form of communication is expensive in terms of latency in the case of cell bouncing, while call blocking or dropping probability reaches as high as 60% for a call rate of 9.5 calls/sec.

Multicast-based Forwarding

Group communication (IP multicast) for the purposes of IP handoffs has been proposed by a number of schemes. In particular, the DAEDALUS effort [42] has employed multicast-based handoffs with packet caching at the neighbouring BSs; a multicast IP address is allocated to each MN by the HA. This approach induces excessive cost in the computation of the multicast tree generated at the HA. The maintenance of such multicast tree at individual BSs, results in considerable multicast state management on a per-MN basis that can render scalability problems as a result of branch join/prune latency and multicast tree reconstruction at the HA for the mobility management scheme [182]. This is because the delivery tree must be recomputed at every new point of attachment of the MN, while the source originates at the HA.

Multicast for Mobile IP handoffs has also been proposed by Seshan [183] and Mysore and Bharghavan in [184]. These schemes are employing IP-Multicast between the MN and its peers (end-to-end) for the addressing and routing of packets to mobile nodes. That is, the schemes require that traffic is forwarded to a multicast group from the CN on an end-to-end basis for the entire duration of MN movement between PoAs.

However, applying multicast for the purposes of handoffs on an end-to-end basis imposes significant computation costs for the multicast tree as described above; the latter is also dependent on the multicast routing protocol used such as PIM-SM [185], CBT [186] or DVMRP [187], as elaborated in [188]. Very recent proposals on multicast for Mobile IP handoffs are also found in [189] with identical problems.

Small group multicast has been proposed by Lee [190] and Ezaki [191] as forwarding mechanisms for MIPv4 during MN's handoff. Both schemes utilise principles from [192] and they employ a multiple unicast destination option at the routing header of a packet. This notion is similar to the route segments of [193] since both schemes rely on unicast routing to deliver the packet. The schemes rely on the provision of *all* CoA destinations to the peer entities from the MN such that packets can be routed in turn towards each access router; this implies that during harsh cell bouncing conditions, this type of virtual small group multicast can miss the receiving MN.

2.5 Conclusions

2.5.1 Handoff Management

From an in-depth analysis of standard as well as most prevalent mobility management proposals we find that simplicity in mobility management protocol design enable scalability during deployment. At the same time, such mobility management designs trade-off simplicity for IP handoff delay performance. This is evident from designs such as Mobile IPv4/IPv6, whereby its signalling appears to be significantly influenced by external delay factors such as end-to-end delay and round trip time variability; this introduces unwanted delay components in the handoff performance of the mobile node (MN).

Also increased signalling overheads arise as a result of simple but persistent location update signalling. In our view, this not a problem since most research indicates that home agents are expected to be well-distributed within a provisioning domain, potentially[24] with one for each serving subnetwork. In addition, the location update signalling incurred by hundreds or even millions of MNs is well over-provisioned at the network core. For different MNs belonging to different subnetworks, mobility management signalling is well distributed within intra-domain routing path segments, isolated from high volume inter-domain mobility signalling traffic.

[24] exception to this possibility is the case where a home agent co-located on an access router serves more than one distinct network interfaces and consequently distinct IP subnets. However, Home agent discovery deals with multi-HA availability over the same link.

As a result, global location management signalling, does not meet the same limitations in IP network infrastructures as is the case in cellular networks; this is because IP networks can massively over-provision the Internet backbone to deal with packet switched signalling that gets eventually distributed to autonomous systems. In addition, autonomous systems support higher routing-path redundancy that can accommodate the routing of such signalling. On the contrary, legacy cellular systems have very little routing redundancy in the face of signalling congestion. At the same time, over-provisioning the cellular network core with similar bandwidth to that of the Internet backbone becomes prohibitively expensive. For this reason, we are confident that micro-mobility mechanisms, although appealing in terms of observed inter-domain locating update savings, will not significantly contribute to IP mobility management performance, other than the reduction of signalling round trip times.

RTT latency savings, however, become a function of the network domain size and the rate of inter-domain mobility. For very large network domains an RTT of 50-150ms is not uncommon. For multi-overlay network infrastructures inter-domain vertical handoffs are bound to be the norm rather than the exception. We have shown that in the case of inter-domain handoff micro-mobility protocols incur *more* signalling than their macro-mobility counterparts.

To leverage RTT variability micro-mobility protocol mechanisms have been proposed. Despite their potential for shorter RTT over control signalling, micro-mobility mechanisms introduce[25] significant complexity in terms of: (i) failure resiliency and (ii) localised mobility agent configuration and distribution over network partitions within an administrative domain. Depending on the individual approach, micro-mobility protocols are expected to incur either:

- extensive changes to the existing IP routing infrastructure
- increased routing state on IP routers in the form of host routes
- additional configuration mechanisms for load balance of traffic over multiple mobility agents (LMAs) distributed at the edges of the network domain
- sub-optimal routing in cases where multiple border routers identify multiple edge routing paths to the Internet backbone. This is incurred by the fact that the MN registers with one LMA at one edge routing path, while traffic arrives to the MN from a different (edge) routing path.

[25]see Annex C.3.

- single points of potential domain-wide protocol failure.

On the contrary, despite their performance trade-offs, simplistic macro-mobility management mechanisms such as Mobile IPv6, can be increasingly scalable with little effect on the network infrastructure. This is clearly *not* the case for micro-mobility protocols. The benefits introduced in the micro-mobility management function of dominant proposals are questionable, since operational scenarios reveal the need for additional mechanisms to sustain localised mobility management savings, originally praised under ideal operational conditions.

The above drive us to conclude that *from a scalability/performance tradeoff perspective, macro-mobility management design strategies maintain an architectural advantage over micro-mobility techniques.* Clearly, this is in line with the doctrine of Fast handoff extensions for Mobile IPv6. FMIPv6 attempts to augment the mobility management function to accommodate provisions for delay seamlessness acceptable for interactive real-time applications. While the proposed extensions are still under investigation, emerging results from independent investigations on the performance of FMIPv6, report prohibitively large handoff delays as a result of an increased number of wireless hosts attached to a wireless link. The reason for this pertains clearly to the dependency of FMIPv6 signalling on access contention, as well as accuracy of unicast tunnel setup, before signals can be propagated between previous and new ARs.

2.5.2 Flow Forwarding

In addition from the perspective of forwarding we find that two viable alternatives may exist: (i) forwarding over a unicast tunnel (ii) forwarding to a multicast group. Solutions of multiple tunnels impose replication of forwarded traffic onto multiple paths. For multiple tunnels to remain scalable they have to remain fixed between old and new ARs. This, however, presents the additional complexity of identifying individual flows for the particular mobile node within the fixed tunnel; once the flow is de-tunnelled at the new AR, the packet has a topologically incorrect destination (the previous CoA of the MN valid at the old AR). As a result nested (pair-wise) encapsulation is required for the AR to distinguish the destination MN at the new point of attachment once the outer header has been removed.

With respect to unicast tunnelling at the new CoA, the significant limitation appears to be the accuracy of resolution of the MN's prospective new CoA. In the event that the tunnel is configured to an incorrect MN candidate CoA, MN's traffic will be forwarded to the *wrong* PoA, reinstating the transient black holes - in terms of packet loss - that Mobile IPv6 introduced. On the contrary, forwarding to a multicast group appears to be free of such limitations. However, it becomes apparent that multicast forwarding cannot be effected at the peers of the MN, that is, HA or CN as is the case with the Mysore et al and DAEDALUS proposals; this is because the multicast tree management becomes prohibitively expensive. On the contrary such cost becomes more manageable when the multicast tree is emanated within the network domain visited by the MN. This is the case with the IDMP and M&M proposals.

The fundamental limitation of multicast forwarding when rooted at the *edge* of a network domain, however, appears to be the classic limitation of micro-mobility protocols: a single point of failure can collapse the operation of the mobility management protocol across the entire domain. In cases where the RP is co-located at the edge LMA, failure of that node implies also total *failure* of the (multicast) flow forwarding mechanism. Where the RP is not co-located with the edge LMA then in the case of an LMA failure the flow forwarding mechanism requires an additional LMA discovery and re-election mechanism to direct all traffic to the RP, assuming that all traffic on the failed edge can be successfully routed to the fallback LMA. In any case, in such situations flow forwarding is guaranteed to experienced sub-optimal routing, as well as increased end-to-end delay. The latter can clearly impact the performance of interactive communications between a live MN and its peers in the face of LMA failure.

Both current unicast and multicast forwarding approaches suggest that a *more robust and distributed flow forwarding mechanism is required.* It is imperative that such mechanism can sustain failure of the forwarding node without destructive effects in flow redirection to all MNs in the network domain. We identify and investigate such a mechanism in Chapter 5.

Chapter 3

Performance short-comings of Mobile IPv6

3.1 Introduction

In the previous chapter we presented a critical perspective on the current state-of-art on IPv6 mobility management mechanisms for next generation wireless Internet. Amongst various protocols designs, we presented Mobile IPv6 as the dominant IP Mobility Management (MM) standard and described its strengths and limitations over alternative protocol proposals.

The wide acceptance of Mobile IPv6, and an enormous amount of collaborative engineering effort in the IETF towards its standardisation, has given rise to a multitude of software implementations, confirming both a solid feature base and its viability as the de-facto IP-MM standard.

To confer, thus, on the efficiency of standard MIPv6 handoff performance over interactive real-time IP application services, we conduct a measurements study over an experimental (last-hop) IEEE802.11 wireless LAN network, employing a real-world Mobile IPv6 kernel implementation.

In this chapter we argue that *reactive acquisition of IP 'connectivity' state at the new point of AR attachment, as effected in the current IPv6 Mobility Management standard, is insufficient to address transmission delay seamlessness, during an IP handoff, for interactive IP application services.*

To prove the above hypothesis we engage first in a series of experimental measurements where a single VoIP packet stream is communicated between a mobile node (MN) and its corresponding peer (CN) over moderate background traffic load. We set as our control hypothesis the case where the VoIP flow is communicated between a stationary MN and its peer while residing on its home network; that is, the MN effects no IPv6 handoff to a visited network. The experimental hypothesis, where the MN effects an IPv6 handoff to a visited network, is then investigated. For the sake of completeness of our analysis we devise a second experimental hypothesis for investigation where the MN performs an IPv6 handoff from the visited network back to its home network. We then compare and contrast handoff performance between the unidirectional downstream component of the VoIP for the cases of MN *stationarity* (no MIPv6 handoff) and *roaming* (MIPv6 handoff) to a *visited* network as well as roaming back to the *home* network.

We analyse the latency implications of the derived MIPv6 handoff performance over other forms of state critical to MN's IP connectivity context at the visited network.

Furthermore, the experimental hypothesis investigates the performance of delay, packet loss and jitter experienced in VoIP communications, during, before, during and after MN's IPv6 handoff. Our findings argue that reactive acquisition of 'IP connectivity' state is not only insufficient but prohibitive for IPv6 mobility management services aiming to support interactive IP application services of wireless access networks.

In addition, we present and analyse the effect of the link layer handoff latency onto the total IPv6 handoff delay, as it emerges from measurements derived from the aforementioned series of experiments. To this end, we complement our measurement traces with a brief simulation study that confers on the effect of wireless-link contention (in multi-node associations on a single AP) on the performance of an IPv6 handoff as effected currently under traditional MIPv6 signalling and the reliance on router advertisements for MN movement detection.

Ultimately we review that part of the generality of the results obtained as well as conclusions arrived at in regards to the delay performance in MIPv6 handoffs, is to a higher degree dependent on the network layer and only to a lesser (but not negligible) degree, dependent on the underlying wireless technology.

In this light, conclusions drawn from link-layer performance over a particular wireless medium are not subject to generalisation. Nonetheless they provide an indication of performance trends tracked *proportionally* by the operational trade-offs constituting the design of the air interface at hand.

3.2 Motivation and Problem Description

IP is by definition a connection-less Internet protocol; that is to say, it requires no explicit connection state (such as sockets manifested in upper layers of the protocol stack such as transport). Nevertheless the ability of a host to effect global routability and addressability in the Internet has been associated with the term Internet or more commonly IP *'connectivity'*. Despite the paradox between the connection-less character of IP and the traditional notion of 'connectivity' in IP protocols and for the sake of simplicity, we chose to adopt this misleading association between term and semantics in our analyses; we employ the term *IP connectivity* to refer to the state required by a host to be or remain globally addressable and routable.

Different forms of IP connectivity state allow the host to be globally routable and addressable under different *contexts*. A few of these contexts critical to the ability of the host to effect or perform in packets communications with its peers are:

- IP mobility management
- Quality of service (QoS)
- Authentication, Authorisation and Accounting (AAA)
- Link capabilities

While each of the particular context of signalling yields a different perspective in IP connectivity, operational reality requires simultaneous existence of most (if not all) of these contexts. For instance, for particular application classes, the network is expected to guarantee specific QoS bounds as part of service level agreements between users and ISPs. More important for the commercial viability of Internet application services, before any host can attach and communicate over different wireless network domains, while in transit, the ISP must ensure at the Access Router (AR) level, that the host is authenticated, authorised and billed as per the pricing tariff or subscription policy agreed with the user.

At the same time, to confirm the *'anywhere-anytime-anyhow'* vision of ubiquitous communications [70, 194, 61], a wireless host is required to roam 'freely' between different wireless networks (or network domains[1]).

The semantics of the term 'freely' for the purposes of ubiquitous computing pertains to the notion of transparent (re)configuration; thus, to roam freely over last-hop wireless networks it is imperative that the MN first attains the required IP connectivity state *transparently*. The first and most important context of IP connectivity state is the one of addressing and routing; without an IPv6 address and a default route no Internet host

[1]For the purposes of this study we confine such vision to last-hop wireless infrastructure network domains as opposed to wireless ad-hoc networks [195].

can be either reachable or routable and thus, can effect no communications in any IP application domain.

As seen in Chapter 2, Mobile IPv6 ensures that global reachability is preserved transparently for the wireless host on the move; when roaming between two (or more) last-hop wireless networks, transparent management of IP addressing and routing state, sustains IP communication with its peers promoting the abstraction of IP(v6) mobility in next generation networks.

Transparency, however in IPv6 mobility management has more than one facets; beyond the transparent *mapping* of addressing and routing state available at two different networks, certain IP application domains render *delay* transparency an equally important goal in such mapping; performance concerns arise when the type of communication of a wireless host, while on the move, involves delay-sensitive IPv6 applications. Example of such applications are interactive multimedia such as Voice over IPv6 (VoIPv6) or Video over IPv6 (ViIPv6).

In such application scenarios, the transparent mapping of addressing and routing state supported by MIPv6 on the wireless host, needs to sustain *seamlessly* the same performance measures, for interactive multimedia services, before, during and after an IPv6 handoff. We term this as the *seamlessness principle* for IP mobility management.

We argue that if the last-hop mobile Internet is to support, in a ubiquitous manner, delay-sensitive applications, IP handoff performance, under MIPv6, must ensure that the seamlessness principle is preserved.

3.2.1 Outline

In this chapter we determine if IPv6 Mobility Management, in the form of MIPv6 handoffs, is indeed capable of preserving the seamlessness principle. We identify the performance characteristics that define the scope of seamlessness for the purposes of this study. We then embark on analysing results derived from a set of experimental measurements conducted on Mobile IPv6 handoffs during VoIP communications between two peers.

The rest of this chapter is structured as follows: Section 3.3 describes the internals of a VoIP system as well as factors critical to the performance of voice communications. Such description allows capturing performance metrics of significance for the purposes of presenting and analysing subsequent measurement results. It further allows to define the subsequent scope of our experimental analysis, by identifying key configuration factors that are both essential as well as feasible for the purposes of measurement and analysis.

Section 3.4 presents the component functions of a MIPv6 handoff. Section 3.5 presents briefly the delay composition of the MIPv6 handoff process through an algebraic account of individual delay components.

Section 3.6 presents the measurements environment established as well as the scope of parameters for both Mobile IPv6 and VoIP communications during the experimental tests. Section 3.7 presents experimental results that identify the impact of MIPv6 handoffs on seamless performance of a VoIP conversation over a wireless LAN, in connection to the individual delay components presented in section 3.5. Section 3.8 presents the respective measure of delay variance as a result of a MIPv6 handoff. Section 3.9 summarises our findings and contributing conclusions on our initial hypothesis.

Supplementary results are presented in Annex D.6, on the influence of the wireless medium, and in particular 802.11, over the MIPv6 handoff process. To this end, a simulation analysis is presented on the effect of contention over the efficiency of the router advertisement interval.

3.3 VoIP system and Quality

VoIP refers to voice communications over IP data networks. Its basic element, speech, in the form of phonemic constituents [196] is an analog signal that varies slowly in time; its bandwidth ranges, typically up to 3.6 KHz. The speech signal alternates between periods of voice content, known as *talk-spurts* and periods of silence; these have been shown to be exponentially distributed, with a mean of 352 ms and 648 ms respectively, according to Sriram and Whitt [197].

To transmit such speech signal over a packet-switched network, the waveform of the speech signal is sequentially, *encoded* into a *compressed* bit-stream and subsequently *packetized*, before scheduled for transmission at the sender. At the receiver, packets are *disassembled*, *decompressed*, *decoded* to waveform samples, *error corrected* and then subsequently scheduled for play-out. During an interactive conversation, the participating system user-host alternates between sending and receiving modes of packetized voice streams.

3.3.1 Encoding of the speech signal

There has been a number of coding schemes for voice, proposed by the ITU-T standardisation body [40, 37, 39]. A coding scheme is capable of both encoding or decoding a VoIP packet stream at a constant bit rate, depending on the role the participating host is assuming; for this reason the coding scheme is typically termed in VoIP parlance as a constant bit rate (CBR) *codec.* A number of CBR codecs have received particular attention as well as popularity becoming an industry standard in most VoIP systems. Amongst these, the simplest is the sample-based G.711 standard using Pulse Code Modulation (PCM) algorithm with a constant bit rate of 64 Kbps[2]; ADPCM-based codecs reduce the bandwidth requirement by encoding the difference between current and previous sample period yielding a CBR flow of 32 Kbps (ITU-T G.726) [38]. CELP-based codecs provide more economical bit rates, with G.729 standard at 8 Kbps and G.723.1 at 5.3 and 6.4 Kbps, at the cost of lower quality audible output, added coding complexity as well as encoding delay [40].

During periods of silence, it is intuitive that the packetized speech signal carries no voice content. It follows, that packets carrying no voice content are effectively a waste of bandwidth. Furthermore, it has been shown in [198, 199] that around 30-40% of a voice conversation is silence, with an average duration of 150-180 sec for a VoIP call. To this end, in an effort to exploit such property in a two-way voice conversation, further bandwidth savings can be achieved if the boundaries of voice activity can be detected, through a technique known as Voice Activity Detection (VAD). Under VAD, when no voice activity is detected (i.e. silence), no audio samples are encoded, compressed, packetised and subsequently scheduled for transmission. However, because the absolute measure of voice activity detected induces to the receiver a sense of speech clipping, the VAD algorithm tends to prolong the talkspurt period by an additional time period known as *hangover* time. Brady [200] employed a long hangover during VAD reporting talkspurts and silence periods exponentially distributed with means of 1.2 and 1.8 sec respectively. A review with a discussion on the on/off voice patterns resulting from modern voice coders can be found in [201]. Generally for small hangover time deltas small talkspurts (200-400) and silence periods (500-700) can be achieved respectively.

In the above manner, perceptible speech clipping is prevented [196]. Moreover, during the period of silence, if VAD is actively suppressing silence packets, the receiver may perceive the lack of silence packets (characterised by ambient background noise) as a VoIP call disconnection. To ensure that no loss of intelligibility in the conversation is incurred, the VAD algorithm, at the start of the silence period, instructs -with a few VoIP packets- the receiver to schedule *comfort noise* packets at the play-out buffer in

[2]Given that voice bandwidth is 3.6-4 KHz, the sampling bandwidth doubles (as per the Nyquist principle) at 8 KHz; each sample is represented with an 8-bit sequence, hence the total encoding of the voice bandwidth amounts to 8000 Hz * 8 bits = 64 Kbps.

place of the expected silence packets. This maintains an audible illusion of a connected call during silence while increasing bandwidth savings.

It is intuitive that by suppressing 'silence' packets under VAD, the VoIP stream its bandwidth requirement is transformed into a variable bit rate (VBR) packet flow with upper bound the codec CBR rate. This will be of importance is subsequent measurements, calculations and analysis of VoIP performance findings during a Mobile IPv6 handoff elaborated in following sections.

3.3.2 Packetisation

The encoded speech, once output as a compressed bit-stream, is then packetized into packets[3] of equal size; this encompasses successive attachment of protocol *headers* by the individual layers of the protocol stack, up to and including the physical layer of network interface. For the purposes of this study we adopt the IEEE802.11b Wireless Local Area Network (WLAN) standard, at the link and physical layer interface. We justify the adoption of such interface in subsequent sections.

The Packetisation sequence of a VoIP packet over an IEEE802.11b link is shown in Figure 3.1

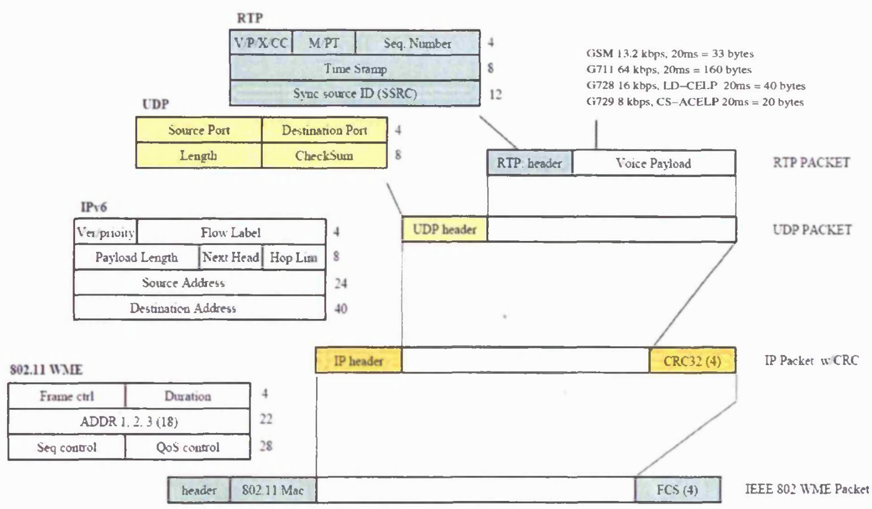


Figure 3.1: Anatomy of a VoIP packet over WLAN IEEE802.11

At the physical layer the frame[4] generated by the link-layer is modulated and transmitted as an analog waveform over the air interface.

[3]the term 'packet' refers to the structure of the bit payload above the network layer of the protocol stack

[4]the term 'frame' refers typically to the link layer (L2) of the protocol stack and follows the same generic 'header-payload' structure -albeit with different syntax and bit format- as the 'packet' abstraction

3.3.3 Delay variance in Received packet flow

Transmission of voice packets over an IP network, is expected to incur variable delay and possibly loss [201], as shown in Figure 3.2. To ensure a smooth play-out at the receiver in the face of delay variability, known as *jitter*, a play-out buffer may be typically used by the *application layer*. In such cases, on receipt, packets are decompressed and decoded, but scheduled for a *later* play-out period. This ensures a continuous play-out and thus, sustains continuity in conversation between sender and the receiver similar to that of a PSTN network.

However, over IPv6 mobility management the IP handoff process[5] may further augment the measure of jitter *since the handoff period is expected to be significantly large*, in addition to the total measure of end-to-end delay between the MN and its CN peer(s). Due to lack of flow forwarding at the previous point of attachment (PoA), packets arriving at that PoA, are forwarded onto the local link whereby the MN is ultimately found to be unavailable (the MN's IP handoff is in progress). Such *black-hole* effect increases packet loss and jitter[6] for MN's received IP flow. Due to the actual loss of packet during the handoff, any play-out technique at the application layer is rendered unable of providing any measure of adjustment in the received packet flow *during* an IPv6 handoff.

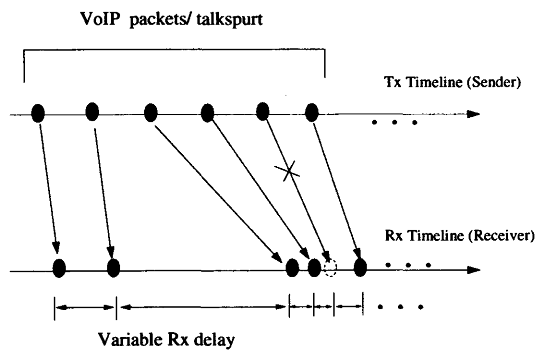


Figure 3.2: Delay variance (jitter) and packet loss during VoIP communications

The potential for play-out adaptation re-emerges when packets can be forwarded towards the MN, during its IP handoff, within the 150-200ms playout deadline; the issue is revisited shortly in Chapter 5. Given that play-out adaptation is an application-level approach not amenable to IP-mobile network-layer control, it extends beyond the scope

[5]MIPv6 does not have a flow forwarding mechanism.
[6]Paradoxically in MIPv6 jitter is manifested as a result of packet loss, as opposed to *late* loss rates due to increased jitter.

of this thesis and thus considered as future research direction.

3.3.4 Decoding and Error Correction

On receipt by the wireless host, the VoIP frame is decapsulated by the MAC layer and submitted in succession to the upper layers of the protocol stack (IP, UDP, RTP). There the RTP voice payload is forwarded to the decoder function of the receiving VoIP system where the speech signal is again reconstructed into phonemic waveforms.

To counteract potential packet loss, decoders may implement, packet loss concealment (PLC) techniques. PLC algorithms attempt to produce a close approximation of the voice sample contained in a lost packet, approaching a perceptually reasonable continuum of the receiving voice output. Simple PLC schemes insert plain silence, background noise or a simply the previously received packet [27]. More sophisticated algorithms establish an approximate voice sample replacement, based on the characteristics of the speech signal, in the neighbourhood of the lost packets; they may employ *interpolation* methods matching the surrounding portion of the lost waveform, or *regenerative* mechanisms by being aware of the codec structure, while exploiting the state of the decoder [26, 202, 202, 203]. A more in-depth analysis of packet-loss recovery mechanisms may be found in [204].

It should be noted that PLC techniques perform satisfactory with loss rates of up to 10% but for small random loss runs (1-3 consequtive packets). Packet loss concealment techniques break down when the loss length approaches the length of a phoneme (up to 100ms), since whole phonemes may be missed by the listener [204].

3.3.5 Quality Impairments in VoIP streams

The quality of voice communication may be affected by a number of impairments that pertain to speech quality or interactivity as well as peripheral delay-related issues.

The quality of packet voice may be affected by low bit-rate compression, well before any transmission over an IP network. In addition, the transmission of voice packets over an IP network is subject to packet loss, while traversing the network fabric, causing degradation in the quality of voice at the receiver. Further loss is incurred in the playout buffer at the receiver, caused by large delay variance as a result of temporal queueing in routers along the path to the receiver. The perceived degradation due to random packet loss can be mitigated by means of PLC at the receiver.

Another aspect of voice communication is the interactivity between the communicating peers, affected by delays arising during transmissions over the IP network. For increased delays, participants perceive 'collisions' during the course of their conversation, which perceptually manifests itself as *simultaneous talking* by both peers. To avoid such collision, the participants resort to conventions of strict alternation between sender and receiver roles in a push-to-talk fashion, as if the connection is half-duplex[7]. This results into highly unintelligible and prolonged conversations.

To achieve a good level of interactivity , the end-to-end[8] delay must remain below the upper bound of 100-150 ms. Longer delays become noticeable by the conversation participants resulting a lower measure of interactivity.

3.3.6 Interactive IP Services over wireless MIPv6

Given the stringent delay requirements for interactive IP application services, we focus on the performance of a packet voice conversation for a wireless host supporting addressing and routing state transparency through the Mobile IPv6 standard.

Under Mobile IPv6 (MIPv6) a wireless host, known as the *mobile node (MN)*, is able to roam transparently between different *points of IP attachment* over last-hop wireless networks.

When the MN changes its point of attachment (PoA), it moves from one IP network to another; a point of attachment may be defined as the entity, typically an Access Router, that enables global reachability to the MN, through some wireless link-layer technology. Such process is known as an IPv6 *handoff.*

Under MIPv6, the MN typically abandons its existing IP network attachment at some random time T_i, before connecting to the new one (assuming use of a single wireless interface) at time T_{i+k}. Thus, there exist some period $T_h = T_{i+k} - T_i$ where connectivity of MN and some point of IP attachment may be *disrupted*; we call this the *handoff period* $T_h = k$. Depending on how the IP mobility management mechanism at hand, handles the handoff process of the MN between the previous and new PoA, it is possible that disruption of IP connectivity may be accompanied by delay and/or packet loss. This is illustrated in figure 3.3.

While many connection-oriented (TCP) application services are designed to cope with intermittent loss of IP connectivity by retransmitting unacknowledged packets, UDP applications - VoIP being one of them - are unable to recover from packet loss, primarily due to the connection-less character of UDP packet transmission.

[7] In a half duplex connection the conversation participant cannot talk and listen to one another at the same time

[8] one-way delay perceived as mouth-to-ear (m2e) latency

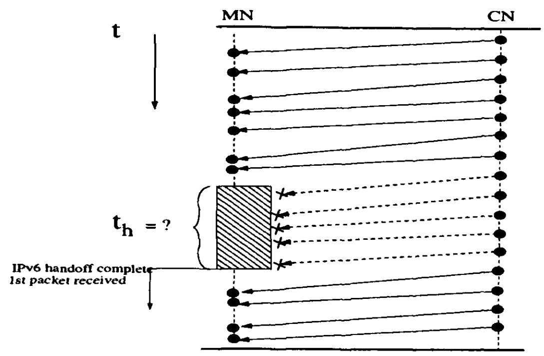


Figure 3.3: Potential disruption of a VoIP conversation during the handoff period

From the above it follows that if VoIP is to be supported for wireless hosts with MIPv6 capabilities, it is imperative that during an IPv6 handoff the seamlessness principle *is* be preserved. This translates to an IPv6 mobility management mechanism, that effects an inseparable embodiment of two qualities in its handoff process, namely:

- incur negligible (or no) packet loss.
- incur low (or negligible) latency

Little or no packet loss incurred during the MIPv6 handoff process, results in a *smooth* IP transition between networks, smooth because transmission of VoIP packets can be sustained during the IPv6 handoff. In that manner, a smooth IPv6 handoff would allow the communication pattern between the MN and correspondent node (CN) to remain unaffected by the handoff process; namely, support VoIP performance comparable to the one of either stationary wireless or wire-line packet voice communications.

Low or negligible latency, incurred by the IP handoff process, results in a *fast* IP transition between networks; this may be achieved if disruption in MN's IP connectivity is minimised while transiting between different IP networks.

Thus, if the mobile Internet is to support, through MIPv6, such class of real-time applications, the IP handoff process must demonstrate these two qualities. If it does not, then additional optimisations and/or design reconsideration to the protocol are bound to be necessary.

It may be seen that the measure of disruption of IP connectivity and subsequently any packet communications during an IPv6 handoff, is clearly dependent on the measure of the T_h. Thus, the mere existence of the handoff period (T_h) gives rise to a number of questions pertaining to MIPv6 communications, interactive or other: (i) how long is T_h ? (ii) can the MN sustain communications during the handoff period? (iii) what are the factors affecting the size of T_h (iv) How is the length of the handoff period (T_h) manifesting itself as an impairment in an interactive real-time communication service such as VoIP?

These questions are receiving the focus of investigation first through an analysis of the MIPv6 handoff process. We determine an analytic measure of the handoff period T_h which we then employ experimentally in identifying the measure of MIPv6 handoff delay during voice communications.

We remind that in there are numerous abbreviations and acronyms used in the area of mobile and wireless IP networking, many of which may be used frequently during the course of our investigation. To aid clarity and conciseness, a glossary of abbreviations and acronyms is therefore provided in Appendix A.

3.4 The MIPv6 handoff process

In this section, we present a comprehensive analysis of the IPv6 handoff process in Mobile IPv6 standard. We first dissect the organic parts of the MIPv6 handoff anatomy and provide algebraic descriptions of the latency characterising their intrinsic functions. We then validate their measure through experimental measurements for different VoIP configuration scenarios.

The MIPv6 protocol enables a MN to roam from its 'home' network to other 'visited' networks while *maintaining* its home network IPv6 address. Such capability is similar to cellular *roaming* in GSM networks [16], whereby a cell phone can be used over different service providers (i.e. cell networks) or over different countries while maintaining its calling number.

Much in the same fashion, the MN remains always addressable by its 'home' IPv6 address; this is the IPv6 address assigned to the MN within its home network. When a MN is away from its home network, packets can still be routed towards it, using MN's home address. As a result, MN movement between networks remains transparent to transport and other higher-layer protocols for application purposes.

Under MIPv6, as the MN *roams* between Internet PoAs, from one IPv6 network to another, it performs the MIPv6 handoff process. Such function is similar to the auto-configuration mechanism employed by a host bootstrapping onto an IPv6 network, but supports some additional functionality:

- the MN must in some manner *detect* fast that it has moved onto a new network.
- during a MIPv6 handoff, transport layer connections remain active; thus the

MIPv6 handoff process must *complete quickly* to minimise disruption from lost or severely delayed packets, or in worst occasions, connection resets.

- once configured, the MN must *inform* its home agent (HA) and each correspondent node (CN) of its new network location in terms of reachability state.

To support the above functions, the MIPV6 handoff process comprises the following event sequence: (i) Movement Detection, (ii) Router discovery, (iii) IPv6 address configuration and duplicate address detection, (iv) CoA Registration and (v) Location Binding Update. The process is illustrated in Figure 3.4. Annex D.1 provides a detailed view of the individual steps of the MIPv6 handoff process together with an algebraic representation of the individual delay components

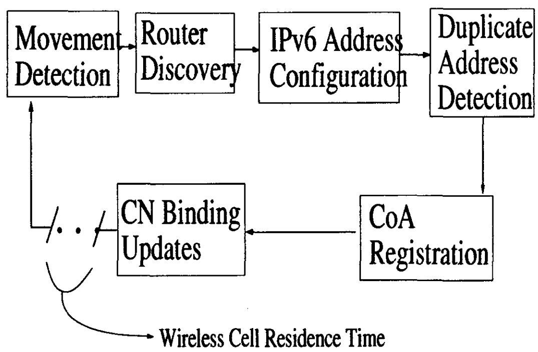


Figure 3.4: Renewal process of the MIPv6 handoff cycle

3.5 Delay components of a MIPv6 handoff

The MIPv6 standard relies fundamentally on core IPv6 protocol functions and in particular the base IPv6 protocol specification [106] and IPv6 Neighbour discovery [107]. Analysing the delay behaviour of a MIPv6 handoff as prescribed by the relevant protocol standards, provides a sound perspective of expected MIPv6 handoff delay performance for validation purposes.

Having identified the constituent steps of the MIPv6 handoff process we now identify the the respective latency components. These are:

- movement detection time (t_d): this is the time required by the MN to detect and establish: (i) a link-layer (L2) handoff and (ii) link-local IP state with the new PoA. This encompasses also the period during which router discovery is completed.

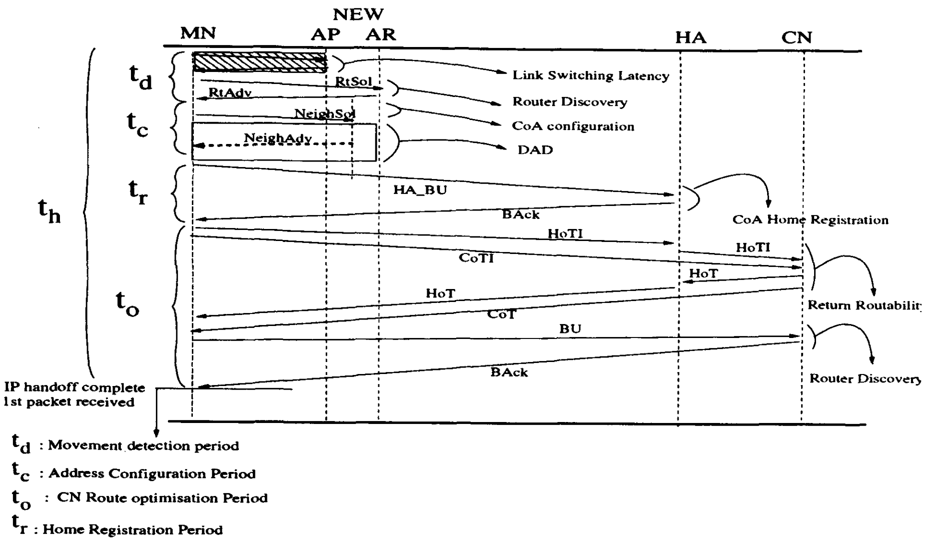


Figure 3.5: IP handoff delay incurred under stateless address auto-configuration by MIPv6 signalling

- IP address configuration time (t_c): this is the period between the establishment of link-local IP connectivity, which is effected statelessly at the MN and the time that a globally routable IPv6 address has been configured. It includes also the period during which duplicate address detection is effected.

- Binding registration time (t_r): past the establishment of context-specific state of the MN, this is the time between the dispatch of a binding update signal to the HA and the receipt of an BU Acknowledgement from its peer.

- Route optimisation time (t_o): this is the period between registering the new CoA with the HA and completing the update of it location bindings with its CN peers. This includes the return routability procedure which, if used, must occur before a BU is sent by the MN to a CN.

The total IP handoff delay may thus be represented by (t_h) defined as the sum of the aforementioned latency components as follows:

$$t_h = t_d + t_c + t_r + t_o \tag{3.1}$$

The individual delay components of the default behaviour of a MIPv6 handoff are shown on the time-line graph of figure 3.5. We note that the graph does not display any other context specific state signalling beyond that of the MIPv6 addressing and routing, during the handoff period. This is because the MIPv6 standard makes currently no explicitly provisions for interaction with any other IP connectivity context (such as

QoS or AAA). We discuss latency incurred by additional state required in different contexts of IP connectivity in section D.6.3.

A detailed analysis of the individual delay components contributing to the total MIPv6 handoff delay is presented in Annex D.2.

3.5.1 Related Work

Surpisingly enough, at the time of writing there have been very few published results on the experimental performance of a MIPv6 handoff.

Perhaps the only work reporting *explicitly* results of MIPv6 handoff performance is the one of Finney and Scott [169, 205]. In this work the authors measured the time taken for a MIPv6 handoff, over a 802.11 Wireless LAN interface. Their experiments employed an early version of the MIPv6 stack with no RR signalling.

Measurements encompassed delay performance over PCMCIA services profiled to log the timimgs of critical points during the MIPv6 handoff process. The authors reported that during a *cold handover*, where the network device was initially totally unconfigured, a total latency of 645ms was incurred with 310ms spent in device driver initialisation[9].

Another 170ms were spent in IPv6 stack activation; such delay was attributed to the architecture of the PCMCIA card services, which employed (slow) user level processes to enable the interface. While active the IPv6 stack required 160ms, to acquire an IPv6 address, followed by binding updates sent to the MN's peers. Round-trip time (RTT) delays were small (5ms) and thus, negligible.

This latency dropped to 165ms under 'warm handoff', where the device is already configured at the link layer, and only dynamic address auto-configuration and binding update transmission was required. Handoff times approaching 5ms were claimed during 'hot handoff', whereby multiple interfaces run in parallel, and care-of addresses could be acquired before the handoff took place.

Their delay figures suggested that allocation of IPv6 addressing state during an IP handoff, generated by itself a delay of 160ms, enough to place any active IPv6 flows on the boundaries of acceptable guarantees for real-time traffic delivery. That experiment involved rather slow hardware[10] running over Mobile IPv6.

The above figures, however, were not reported with some degree of statistical confidence to confer robustly on the performance of MIPv6 handoffs. For instance, the

[9] that is before any network configuration could take place.
[10] P133 laptop equipped with 10Mbps Ethernet PCMCIA interfaces

authors provide no explicit analysis on the influence of duplicate address detection in the MIPv6 handoff process. Furthermore, since the date these experiments were conducted the MIPv6 specification has changed significantly.

Thus, the above results may only provide some very coarse measure on MIPv6 handoff performance. Such measure would be insufficient for assessing the efficiency of a MIPv6 handoff in support of interactive real-time IP services. To this end, we conduct a detailed series of experimental measurements on IP handoff delay performance, over the MIPv6-enabled wireless networks.

3.6 Experimental Measurements on MIPv6 handoff delay performance

To identify concretely the true IPv6 handoff latency incurred by the MIPv6 standard we perform a series of experimental measurements on a real-world software implementation. This is the Linux HUT MIPv6 implementation [206].

3.6.1 Measurements Environment and Methodology

Experimental measurements conducted on this MIPv6 implementation carry significant value in regards to the establishment of a statistically credible performance metric for MIPV6 handoff delay performance. The value of such measurement, is multi-fold:

- The HUT implementation of MIPv6 provides and objective view of IPv6 handoff performance under MIPv6, as it is implemented by an independent party, while its code-base remains open to public scrutiny (open source).

- The HUT MIPv6 code-base is a well-known and accurate implementation of the MIPv6 specification. It has been used consistently and repeatedly by a number of research projects or institutions. Thus, a set of experimental results derived from this implementation would provide a sound set of initial assumptions towards subsequent valid inferences in line with the modus ponens principle of scientific investigation.

- experimental results derived from this implementation remain replicatable by other independent research bodies by adopting a similar control hypothesis.

For simplicity we focus our trace measurement analysis to a unidirectional component - downstream in particular - of the typical two-way VoIP packet stream, assuming a *fixed* CN. Furthermore, the communicated flow is assumed to be a constant bit rate flow with no silence suppresion. The case where: (i) the CN is also a mobile node and (ii) a variable bit rate flow is communicated by way of applying silence suppression on the communicated talkspurt stream, is considered out of scope in this investigation, and thus left as a future research direction.

The above experimental focus does not detract from the validity or generality of the derived results. This is because for the upstream component of the VoIP session, the CN experiences a nearly[11] *identical* measure of MIPv6 handoff delay, assuming a constant bit rate flow in both directions of the VoIP session.

For a VoIP session where silence suppression is enforced in both directions, we may *approximate* MIPv6 handoff delay performance by considering characteristic measures of talkspurt (on) and silence (off) periods from actual VoIP session. We recall from Section 3.3 these periods may be modelled as exponentially distributed periods with mean 352ms (talkspurt) and 648ms (silence) respectively. Statistically these mean values may also represent a probability of 0.35 of voice content and 0.65 of silence during the period of MIPv6 handoff delay. By looking the performance of the talkspurt probability within, or at the boundaries of the meausure of MIPv6 handoff delay we can derive an expected measure of MIPv6 handoff delay affecting VAD-assisted VoIP sessions. These figures can be validated by future experimental measurements focusing specifically on VAD-assisted VoIP communications over MIPv6-enabled wireless LANs.

Experimental Set

For the purposes of IP handoff measurements over MIPv6 we have devised the (`Mobirig6`) experimental testbed, shown in figure 3.6. The testbed comprises of two Wireless LAN cells, each serving a different IPv6 sub-net through an Access Router (AR). Communications over the air interface were effected with WEP encryption.

The first AR is configured to also serve as a Home Agent (HA) for the Mobile Node (MN), while the second employs plain IPv6 routing (MIPv6-unaware). Implementation improvements on the frequency of Router advertisements provided by the Monash Research group [207] allows configuration of the average Router Advertisement interval to 50 ms, together with optimisation options that may be exploited by the MN; in particular, the MN is configured to effect a router solicitation every 250ms (as per the ND specification) for the purposes of movement detection. Table 3.1 provides the configuration variables and settings that comprised the default behaviour of MIPv6.

[11]Nearly because the upstream path is not guaranteed to have a measure of one-way delay equal to the downstream one.

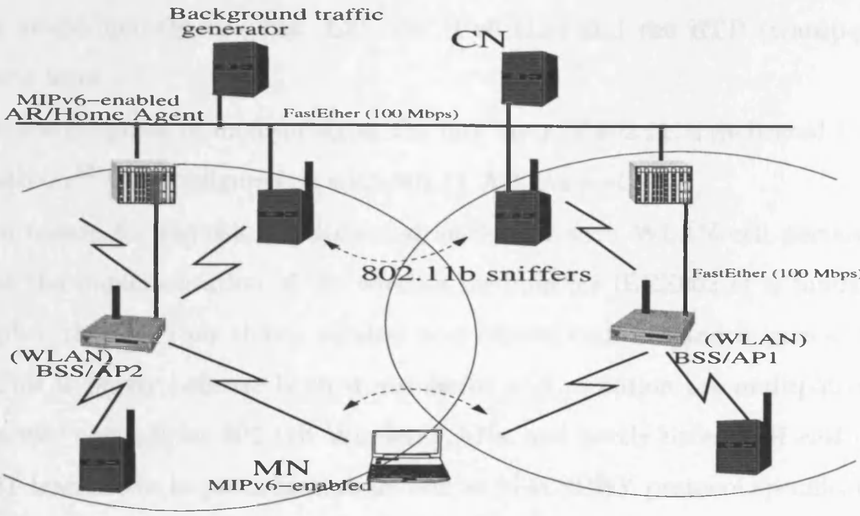


Figure 3.6: Mobirig6 experimental testbed

RtAdv Configuration Parameter	value (sec)
RTR_SOLICITATION_INTERVAL (MN)	0.25
RTR_SOLICITATION_MAX_SENDTIME (MN)	5
MinRtrAdvInterval (AR)	0.03
MaxRtrAdvInterval (AR)	0.07
AdvIntervalOpt (AR)	yes
AdvSourceLLAddress (AR)	yes
AdvValidLifetime (AR)	600
AdvPreferredLifetime (AR)	600

Table 3.1: Configuration variables effecting default behaviour of Mobile IPv6

The testbed hosts further a MIPv6-enabled Corresponding Node (CN) and Mobile Node (MN); the CN remains fixed, configured on a wire-line Fast Ethernet interface, while the MN remains mobile, configured over an IEEE802.11 Wireless LAN interface, associating with each of the two 802.11 APs, in sequence.

We may note that for the purposes of our experiments and for the scale of the experimental setup, configuration of additional visited networks would not offer any additional insights with respect to MIPv6 handoff performance. The reason for this is the fact that (i) the RTT between home and visited networks is fixed and nearly negligible and hence would not influence the magnitude of handoff latency significantly, (ii) the behaviour of a MIPv6 handoff during the transition between two visited networks can be emulated with the handoff performed between the home and a visited network, so long as no *residue* addressing or routing state remains in the MN or ARs, during the next MIPv6 handoff renewal period.

The testbed configuration encompasses monitoring at 3 separate layers of the network stack, namely the link (L2), the IPv6 (L3) and the RTP (transport/meta-transport) layer.

For the purposes of monitoring at the link layer of 802.11, a dedicated frame capture/analyser[12] was configured at each 802.11 AP *channel*.

The reason for requiring a dedicated sniffer for each WLAN cell pertains to the fact that the implementation of the wireless medium for IEEE802.11 is fundamentally half-duplex; that is to say that a wireless host cannot transmit and receive at the same time. This is partly because both transmission and reception are multiplexed on the same *carrier channel* for 802.11b Wireless LANs, and partly because of *cost efficiency* in 802.11 transceiver implementation as well as MAC/PHY protocol specification.

As a result, during communications the wireless MN cannot monitor MAC control traffic over the 802.11 link, while engaged in IP communications with the CN. This implies also that, any packet capture effected on the same 802.11b MN interface monitors only traffic destined to MN's IPv6 (or higher) layer.

For the purposes of VoIPv6 communications we employed the RATv6 (MN) [208] and KPhone (CN) [209], communicating through a GSM encoding with a bit rate of 13.2 Kbps. The encoding employs a packetisation rate of 50 packets/sec dispatching voice payload every 20ms.

MIPv6-enabled communications are effected over an average offered load of 350 Kbps over each wireless link towards a stationary wireless host. It is noted that at a signalling rate of 1 Mbps the maximum offered throughput load before saturation is found to be around 700 Kbps [51, 210]. The traffic load comprized of four unidirectional flows: two TCP flows[13] each at 15 Kb/sec and a UDP flow at 10 Kb/sec all destined to the same stationary wireless node in each WLAN cell. The measured VoIP flow consumed 1.62 Kb/sec.

802.11 AP configuration

Each dedicated WLAN frame analyser monitors a single 'channel' as the per 802.11b specification [17]. As elaborated in Section D.6.1, the term 'channel' does *not* refer to a discrete, single frequency band. This is because the spread spectrum character of signal modulation. The spreading implies that the actual RF signal energy is not constrained withing a single discrete frequency; instead it is spread over a small frequency range.

With respect to the data rate on each AP, we configure both networks to support a

[12] AiroPeek NX is a commercial high-resolution IEEE802.11 trace analyser provided kindly by Wild-Packets Inc. for the purposes of this study

[13] Each tcp flow was an FTP file of 200Kb size and a UDP flow of 100Kb.

signalling rate of 1 Mbps. This is because the MN is expected to experience the lowest signalling rate at the coverage boundaries before the IP handoff is initiated.

Experimental Metrics

Analysis of the measurement traces employed the following metrics:

1. MIPv6 handoff latency : defined as the absolute number of packets lost during a MIPv6 IP handoff.

2. MIPv6 handoff-induced packet loss: defined as the number of packets lost during a MIPv6 handoff.

3. jitter: defined as measure of delay variance. The mechanism for jitter calculation is presented in Annex D.3.

Perceptible IP handoff latency may then be defined as the perceptible number of packets lost during a MIPv6 IP handoff. We assume from human factors in real-time multimedia applications an upper limit of 200ms for one way delay. It is important to note that, the respective number of packets lost during such period is dependent on the packetisation rate and/or whether functions like VAD are available on the encoding source.

3.7 Experimental Results

In the following section, we present experimental results obtained from a VoIP communication flow between the MN and its CN peer. We break the bidirectional VoIP stream and focus on the analysis of the unidirectional flow sent by the CN downstream to the MN. In this series of experiments we effect 20 MIPv6 handoffs, where each handoff is effected by the MN by transiting from its home network to the visited network and vice versa. Since we analyse the behaviour of the MIPv6 handoff alone, we employ no mobility pattern to simulate a realistic mobility pattern for the MN. Instead, handoffs are effected periodically every 10 sec.

We may note that the RTT delay between the MN and its CN peers was on average 2.3 ms during transmissions at a signalling rate of 1 Mbps; hence, one-way delay was considered constant and thus assumed to be negligible in comparison to the total measure of MIPv6 handoff delay.

IPv6 layer

During a MIPv6 handoff we observed that the MIPv6 network stack does not remove the IPv6 CoA configured during its previous IP handoff to the visited network. Furthermore, the MIPv6 stack at the MN is normally *unable* to clear the entries augmenting the Neighbour Cache of the visited network's AR. This is important for the purposes of measuring accurately the true MIPv6 handoff delay incurred during the handoff; during a MIPV6 handoff the new AR has typically no IPv6 connectivity state for the MN with respect to its addressing and reachability. Hence, *any Neighbour Cache entry residues at the visited network AR, or a residue CoA configured from a previous MIPv6 handoff would not provide the correct MIPv6 handoff delay as per the MIPv6 specification.*

To this end, the `Mobirig6` measurement testbed enforced a mechanism that ensures that *no CoA at the MN or Neighbour Cache entry residues affect the measure of MIPv6 handoff delay during experimental measurements.* This is achieved by explicitly 'flushing' both the CoA on the MN as well as all neighbour cache entries on the visited AR upon return to the home network. On the contrary, the co-located HA/AR system does not have its neighbour cache flushed as it defends the home address of the MN.

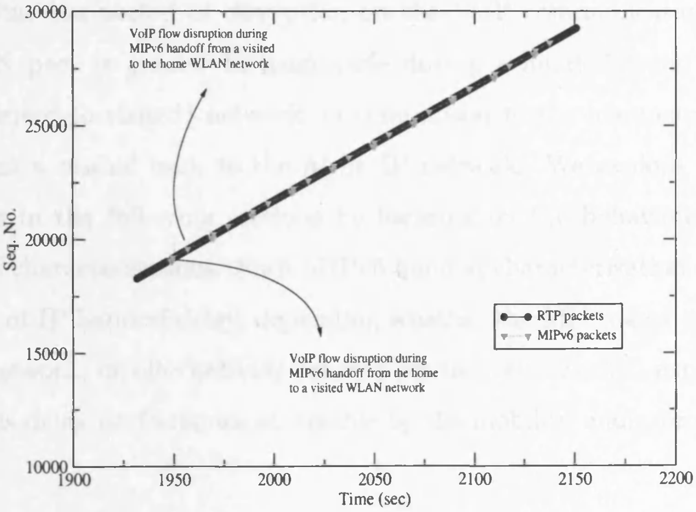


Figure 3.7: Disruptions on VoIP communications for the MIPv6 mobile node, while in transit, during an IP handoff to a visited network or upon its return back to the home network

Figure 3.7 presents a birds-eye view of the effect of MIPv6 handoffs for the wireless host on the move, while a VoIP session is in progress. From the unidirectional flow sent downstream towards the MN, we observe that during the handoff between the home and a visited network, two kinds of VoIP disruption arise: (i) a *dominant* one occurring during MN's IP handoff *from the home to the visited network* and (ii) a *quasi-quiescent* one, occurring during the MIPv6 handoff *from a visited to the home network.*

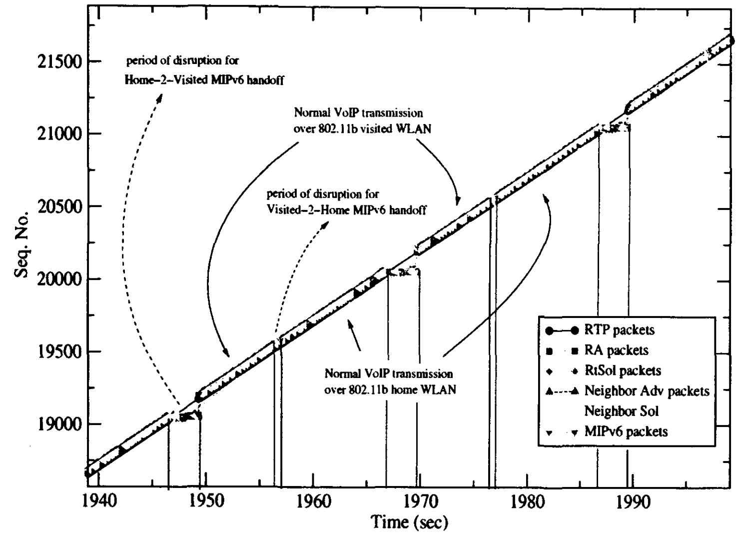


Figure 3.8: Close-up of VoIP flow disruption during a MIPv6 handoff to a visited or the home IP network.

The disruption becomes clearer as we close-up onto the trace in Figure 3.8; we may observe that the period of disruption on the VoIP communication between MN and the its CN peer is *greater* in magnitude during a handoff from the *home* to a *visited* (or a visited-to-visited) network, in comparison to the one incurred during the IP handoff from a *visited* back to the *home* IP network. We explore the reasons for such behaviour in the following sections by focusing on the behaviour of the two of MIPv6 handoff characterisations. Such MIPv6 handoff characterisation emphasises: (i) the magnitude of IP handoff delay, depending whether the MN roams *away* or *towards* it own home network, or alternatively *between* visited networks (ii) exposes important insights towards delay performance attainable by the mobility management mechanism at hand.

Roaming to a visited Network

Figure 3.9 presents a close-up view of the MIPv6 handoff process and dominant delay component as the MN transits from its home to a newly visited network. We observe that after the last VoIP packet is received (#51904), the packet capture at the MN experiences a period of silence of about 378ms before the first router advertisement is received. During this period of silence only a router solicitation is 'voiced' from the MN 182ms after the receipt of the last VoIP packet, in search of a router advertisement from its current point of attachment. The period of silence in VoIP transmissions is due to link-switching, (commonly known as *L2-handoff*), between 802.11b APs is analysed statistically in Section D.6.1.

A router solicitation is sent by the MN if it has not received a router advertisement within a period equal to 3 times the existing router advertisement interval on that link. This aims towards a faster detection of IPv6 network movement (change of AR attachment) for the purposes of IPv6 mobility management. Without this the MN would rely solely on the lifetime of a router advertisement, which if long, would result into a slow detection of IP movement, as the RtAdv lifetime would typically expire much later than the IP-movement of the MN.

Handoff delay and IPv6 movement detection Lack of periodic router advertisements causes the MN to assume *erroneously* that the IPv6 link is still available (perhaps on a different advertising interface). To this end, the MN sends a router solicitation out on its wireless link, but receives no immediate router advertisement. The reason for the latter is two-fold, given that the semantics of router solicitation in IPv6 are overloaded with the additional process of network prefix switching as a result of network IPv6 mobility.

The first reason pertains to fact that the MN is actually undergoing a L2-handoff during the MIPv6 handoff between adjacent IPv6 cells, while on the move. Secondly, unicast RtAdv responses to router solicitation are, by way of the Neighbour Discovery specification, delayed randomly between 0-500ms; such delay aims to prevent multiple on-link routers from simultaneously transmitting to soliciting IPv6 hosts [107].

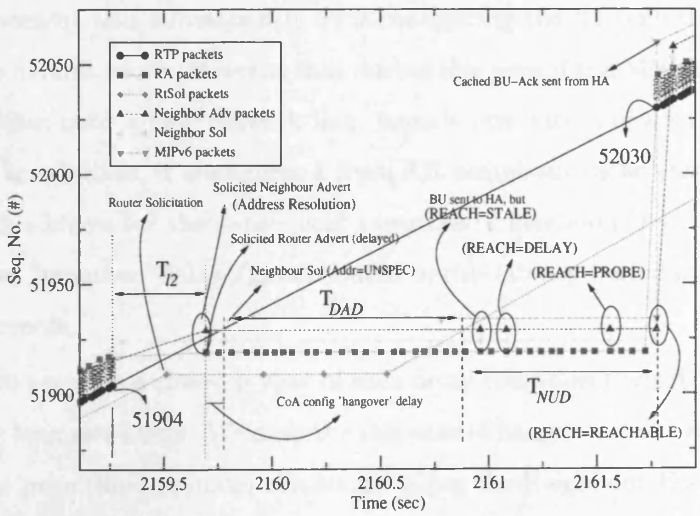


Figure 3.9: Protocol behaviour of IPv6, MIPv6 and RTP layers during a MIPv6 Handoff returning to the home network at 1 Mbps in both previous and new 802.11b AP

Completion of the L2-handoff, *may* be followed by a subsequent unicast RtAdv response to the initial router solicitation. Such response is, however, not guaranteed; this is dependent on the amount of allowed MAC retransmissions by the link-layer on that particular frame. It is, thus, possible that the RtSol message may not be received by the AR if the number of MAC frame (MPDU) re-transmissions is exhausted well before the period of completion of link-layer switching. This is dependent both on the configuration of allowed MAC retransmissions on the MN and/or the contention in effect on the air interface.

In the event that the router solicitation *is* responded to by the AR, the unicast RtAdv message is preceded by address resolution, since the AR must first resolve the link-layer address of the MN attaching to this network (once link-layer switching has completed). This is important for the purpose of neighbour reachability (NUD), manifested by subsequent solicited neighbour Advertisements to that MN.

Address resolution is effected by sending a Neighbour Solicitation message to the MN's solicited-node multicast address for its link-local address. The MN responds with its link-layer address sent through a Neighbour Advertisement back to the AR's link-local address. Upon completion of address resolution the MN receives the solicited router advertisement. Both address resolution and solicited router advertisement complete within 5.23ms after the completion of the L2-handoff between APs.

Receipt of the solicited RtAdv message experiences a *hangover* delay until the initiation of the DAD process. During this delay period, the MN is processing the router advertisement and subsequently auto-configuring the its tentative CoA as well as updating its default route. It seems that during this period the MN actually 'realises' an IPv6 transition onto a new network link, namely one with a new subnet prefix and default route. In addition, it configures a fresh AR neighbour cache entry, with respect to the AR's L2 address for the subsequent purposes of neighbour reachability (NUD). Hence, the term 'hangover' delay T_{ho} attributed to this latency period prior to initiation of the DAD process.

Figure 3.10 presents a close-up view of such delay component which for this handoff sample lasts as long as 84.9ms. We describe this case of hangover delay as *nominal* since it describes the prescribed protocol behaviour as per the Neighbour Discovery protocol modified by the MIPv6 standard. It becomes obvious that the nominal hangover delay describes also *the race condition* between the processing of the solicited router advertisement and a router advert arriving periodically (potentially as soon as the former) through the all-node multicast group.

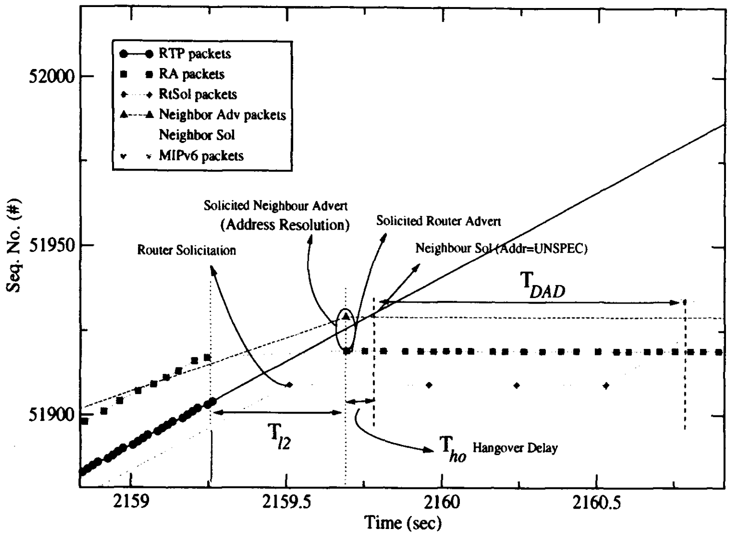


Figure 3.10: Hangover delay component resulting after receipt of RtAdv and prior to DAD initiation through neighbour solicitation

What is interesting, however, with respect to movement detection is that the nominal hangover delay did not describe the majority of the total number of handoffs recorded during the experiments. Out of a 1000 handoffs[14] performed, only a small portion presented the above protocol behaviour.

In particular, only 178 out of 1000 total handoff exhibited the *expected* IPv6 movement detection behaviour, as prescribed by the Neighbour Discovery [107] standard. This represents about 17.8% of the total MIPv6 handoffs measured; the remaining set of MIPv6 handoffs (82.2%) was distributed with a ratio of 0.6232 and 0.3768 respectively, between two forms of movement detection: (i) a MIPv6 handoff whereby the router solicitation is lost during L2-handoff time; (ii) a MIPv6 handoff whereby *no* router solicitation is lost during the movement detection process.

Figure 3.11 demonstrates the case (i) of movement detection, representing the 51.22% of MIPv6 handoffs away from the home network. It may be seen that the router solicitation is rendered lost as a result of the L2-handoff and after exhausting the number of preset MAC frame retransmissions. We identify this as a sap[15] form of movement detection and specifically as *sap-reactive movement detection* and the respective delay as *sap-reactive hangover delay.*

During measurements, the number of MAC retransmission was set to 5 given that the mean inter-frame retransmission period at 1 Mbps emulating cell boundary condi-

[14]performed in 100 iterations of 10 handoff runs - where a single handoff is counted as both home-to-visited and visited-to-home. This is because the MN would 'hang' the RAT VoIPv6 application after the 15th handoff consistently throughout our measurement trials.

[15]implying very slow response irrespective of the external control signalling stimuli.

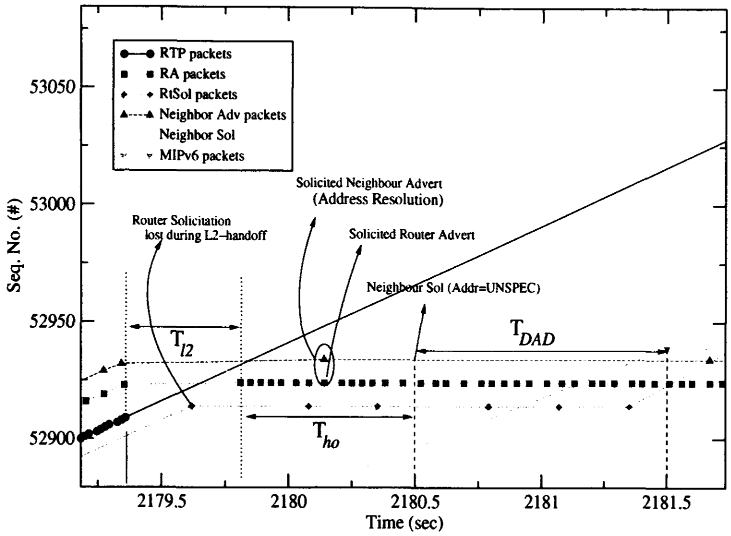


Figure 3.11: Hangover delay caused by lost Router Solicitation manifested as temporal insensitivity to Router Advertisements

tions was found to be 1.23ms with std. Dev. 0.001345. A larger number of retransmissions was not considered to avoid increasing contention during retransmissions with a cascading effect on the completion of the L2-handoff. The particular sample instance of lost router solicitation rendered a hangover delay of 687.9ms. We elaborate further on the effect of MAC retransmissions in Section D.6.1, where we analyse the influence of an L2-handoff.

It is interesting to note two unexpected events arising as a result of the loss of a router solicitation during the L2 handoff: (i) the MN remains insensitive to periodic RtAdv messages sent to the all-nodes multicast group; (ii) router solicitations continue to be transmitted by the MN even after the dispatch of the Neighbour solicitation signifying the start of the DAD process.

In the second form of movement detection identified, accounting for about the 30.98% of the total number of measured handoffs, the hangover period commences from the receipt of the delayed (due to L2-handoff) solicited RtAdv message, without loss of the router solicitation. A sample instance of this case is shown in figure 3.12 where the hangover delay amounts to 355.2ms, during which the MN solicits yet another router advertisment.

This form of detection of IP movement is identified as *numb-reactive movement detection* and the respective delay as *numb-reactive hangover delay.* This action on the part of the MN may only be justified by the possibility that the previous solicited

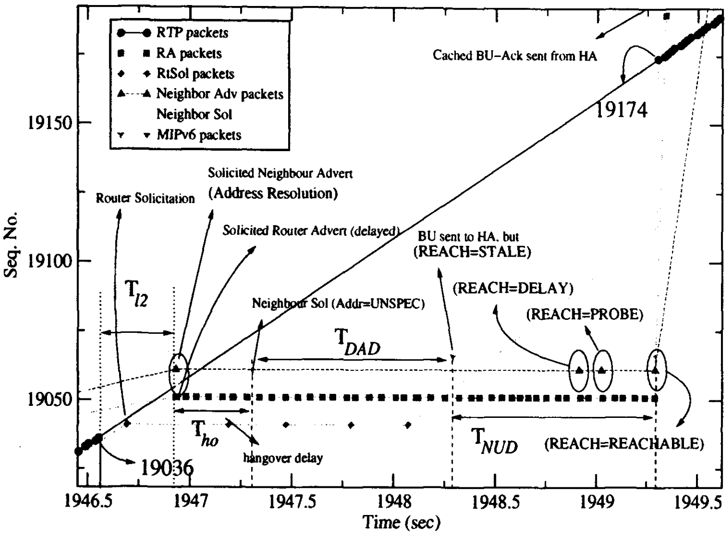


Figure 3.12: Hangover delay caused by delayed solicited RtAdv, emerging as temporal insensitivity to Router Advertisements

router advertisement was not utilised for the purposes of address configuration. Again during this form of movement detection, the MN *remained insensitive to periodic RtAdv messages.*

From a protocol-behaviour perspective there exists no clear reason that justifies such magnitude of hangover delay for both these cases, where the MN remains insensitive to periodic RtAdv messages. We cannot exclude the possibility of inefficiencies in the particular kernel implementation, although as we see further, during the MIPv6 handoff back to the home network, response to periodic RtAdv messages is consistent as per the ND protocol standard, for all 1000 handoffs measured. We have seen, however, in Section 3.7 that improved performance on the MIPv6 handoff back to home network may attributed to the co-location of the HA function onto the AR device.

We have excluded the possibility that the MN delays in joining the all-nodes multicast group (ff02::1) since the measurement trace captures correctly RtAdv messages destined to that group in all cases of promiscuous, non-promiscuous or external tracing[16] modes, with no change in the MIPv6 handoff behaviour.

Nevertheless, what remains clear, in the face of *hangover* delay, is that the MN relies solely on solicited the unicast RtAdv message sent by the visited AR. Hence, *the timing of transmission of a Router solicitation by the MN, the interval between successive router solicitations, or any delay by the AR in responding with a unicast*

[16] by a separate wireless host

RtAdv message back to the MN are of prime importance to the task IPv6 movement detection and subsequently MIPv6 handoff during VoIP communications.

Handoff Delay and DAD On receipt of the second solicited router advertisement, the MN proceeds to complete its interface configuration by configuring a CoA through the stateless address configuration function. We observe that upon receipt of the solicited RtAdv signal until the NeighSol signal, signifying the start of the DAD function, there exists a delay of 78.94ms (see figure 3.12).

This may be justified as the total processing time of the between the receipt of the router advertisement until the dispatch of the NeighSol packet that checks the uniqueness of the CoA. This delay component must also include both the address configuration time (T_{EUI64}) as well as update ($T_{RouteUpdate}$) of the MN's default route. Address configuration is complete on average after 1000.2ms with the MN dispatching immediately the first BU message to its HA.

Handoff Delay and Neighbour Reachability Following completion of the DAD process, it is interesting to note that while upstream reachability seems to have been resolved by means of the initial solicited neighbour advertisement by the MN, downstream reachability remains at the STALE state at the neighbour cache of the new AR; although the RTT between MN and CN is only 4.2ms it takes nearly another 1000ms before downstream reachability is restored for the MN at the AR.

The above is also attested by the 3 subsequent solicited neighbour adverts whereby the AR proceeds through reachability states (`DELAY` and `PROBE` by soliciting a NeighAdv message from the AR. Finally, upon solicitation of NeighAdv from the AR to the MN, the latter responding with its CoA as its source address, sets the reachability state at the Neighbour Cache of the AR to `REACHABLE`. This signal concludes the setup of reachability state with subsequent delivery of the BUAck response sent previously by the HA. We note that past the DELAY state the MN is probed with neighbour solicitations sent every RETRANS_TIME interval (1000ms) until the solicitations are responded to with a neighbour advertisement.

Such measure of latency (T_{NUD}) has not been reported in the past either explicitly or implicitly as a possibility of induced latency, despite the fact that the Neighbour discovery protocol signifies such possibility during the process of Neighbour Unreachability detection.

The former effectively completes the CoA registration process, while the MN proceeds to effect the appropriate route optimisation by updating its bindings directly with its CN peers. Such a process is accompanied by the first VoIP packet arriving through IPinIP [211] encapsulation only 9.1ms after. This is followed by the completion of the RR process whereby the MN dispatches both HoTI and CoTI signal to the CN as illustrated in Figure 3.5. Without accounting for the variability in delay incurred by the RTT of the paths between MN-HA, HA-CN and MN-CN, we observe 6ms and 7.24ms as delay of the HoTI and CoTI message sequences respectively. It is interesting to note that although the HoTI sequence would travel a longer path since it would reach the CN through HA, the measurement shows that it is the CoTI message sequence that took longer to complete. This is surprising since all machines are of the same processing capacity with the CoTI sequence expected to experience a shorter path and thus a faster completion. Nevertheless, it is more likely that in the presence of RTT variability the aforementioned difference would be uninteresting and thus negligible for the purposes of total handoff delay.

Returning back at the home network Latency of a MIPv6 handoff during the return of the MN back to its home network observes a different behaviour. In particular, just before the initiation of the MIPv6 handoff at the MN, the last VoIP packet is followed by the usual neighbour solicitation by the MN, checking for upstream reachability before the transmission of the next VoIP packet.

The drop of the signal-to-noise ratio (SNR) below the expected BER threshold initiates the normal link-layer switching for a period during which the MN does not receive any router advertisements. During this time, the threshold of $3 * RtAdvInterval$ is exceeded and thus, the MN sends a router solicitation in search of a RtAdv to check its AR attachment. This is effected about 200ms after the last NeighSol message sent out to refresh reachability with the (previous) AR attached at the time.

Interestingly, the delay of 200ms has not effected a complete link-switching between the two neighbouring APs defining the coverage area of the visited and home network; instead, MN's link layer actively *scanned* the air interface for the stronger SNR available by the neighbouring APs, before the link-layer handoff was effected.

Having completed this scan process, the MN returned momentarily back to the associating AP, whereby after receiving 3 subsequent VoIP packets (sent to the MNs CoA with a type-2 routing header) it effects its link-layer handoff decision by attaching back to the home WLAN network (since that was the AP with the dominating SNR). This is first experienced by receiving a router advertisement from its home AR.

Receipt of the first router advertisement (carrying the HA information option) unlike the behaviour experienced at the visited network, results into a neighbour solicitation to the solicited-node multicast address for its home address (with source the unspecified address); this effectively implies initiation of the DAD process for its home address. For the purposes of (also) address resolution, a second neighbour solicitation is sent out to the solicited-node multicast address for the address of its HA. Since the MN's home address is 'defended' by the home agent the DAD-oriented neighbour solicitation is responded immediately (within 8.86ms) with a neighbour advertisement by the HA, signifying to the MN that its home address is in use by the HA.

This event informs the MN that it has returned back to the home network; within less than 1ms it sends a BU message to the HA, which in turn within 3.9ms responds with BUAck. The MN proceeds instantly (less than 1ms) to inform all nodes on the link by issuing a neighbour advert such that, nodes on-link update the respective neighbour cache entries for both its link-local and global home address to map onto the MN's link-layer address, instead of the HA's one.

This is followed by HoTI message send immediately towards the CN (through the HA), which is responded by the CN after 8.05ms (through the HA) back to the MN. This is followed by an immediate BU sent directly to the CN, which responds with the next available VoIP packet sent directly to the home address of the MN. At this point the MIPv6 handoff process for the MN returning home, is complete and hence, VoIP transmissions continue over the home network.

It is interesting to note that during the MIPv6 handoff the delay incurred is significantly lower than the delay experienced when a handoff is performed towards a visited network. This is accounted by two reasons: (i) the DAD process resolves immediately without a need to resort to expiry of the RETRANS_TIME interval after which DAD is either successful or fails (ii) there exists no delay in exchange of reachability information with respect to neighbour unreachability detection.

The significant difference in IP handoff performance between a *home-to-visited* and a *visited-to-home* MIPv6 handoff is two-fold: (i) there exists an entity on the home network that *'defends'* the home address of the MN, should any IPv6 traffic be directed to the latter, while away from the home network; (ii) the HA mobility management entity is *co-located* on the AR of the home WLAN network.

With respect to (i), such arrangement allows both address resolution and duplicate address detection to be acted upon *instantly* during a neighbour solicitation. This is

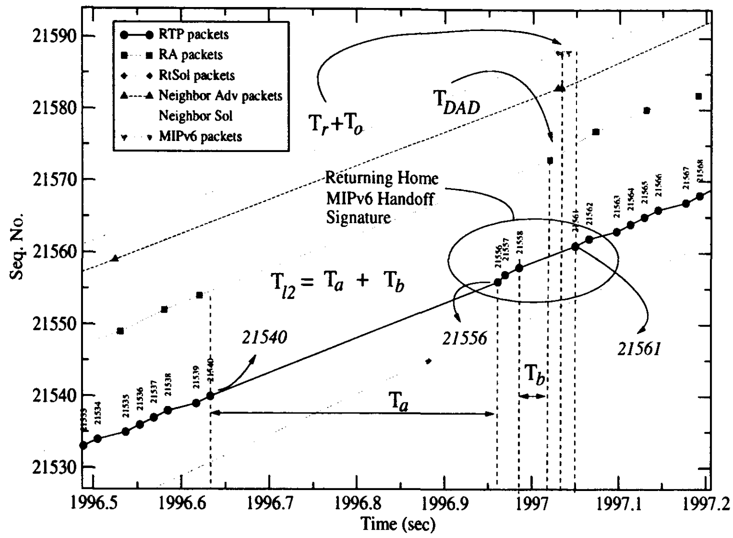


Figure 3.13: Protocol behaviour of IPv6, MIPv6 and RTP layers during a MIPv6 Handoff returning to the home network at 1 Mbps in both previous and new 802.11b AP

clearly validated by the immediate response on a neighbour solicitation for the purposes of DAD resolution for the MN while returning to its home network as seen in figure 3.13. This implies that *it is the existence of an additional function at the home network that allows for fast resolution of the DAD process during a visited-to-home MIPv6 handoff.*

In regards to (ii) neighbour reachability information for both HA and the enabling (for the MN) AR is co-located under the same neighbour cache. Given that HA deals explicitly with reachability information pertaining to the IPv6 Mobility Management, the (access) routing function on the this host simply exploits bidirectional reachability information already harvested by the HA. That is to say, co-location of the HA onto the AR is the cause of the apparent improvement in MIPv6 handoff delay performance at the home AR with respect to neighbour unreachability detection.

To validate the above claim we separated the HA from the AR entity by placing them onto two different hosts, on the same home network link. When a visited-to-home MIPv6 handoff was effected under this network configuration, the T_{NUD} delay component re-emerged in an identical manner to the one presented in the home-to-visited MIPV6 handoff case shown in figure 3.9. This confirms our hypothesis, that co-location of the HA onto the home network AR influences significantly the delay imposed by neighbour reachability resolution between the AR and the MN. *If the HA is not co-located with the AR function onto the same host, then the visited-to-home MIPv6 handoff will also experience the T_{NUD} delay component.*

3.7.1 MIPv6 handoff performance statistics

To attain a reasonable measure of statistical confidence in the aforementioned delays observed during a MIPv6 handoff away or returning to the home network, we repeated the measurement process for about 1000 handoffs and collected traces which were subsequently analysed statistically off-line. In particular, we approximated our measurement dataset onto a statistical distribution that best describes the behaviour of this dataset for the particular component of MIPv6 handoff delay.

To this end, we conducted a set of goodness-of-fit (GoF) tests, that measure how well the sample data fit a ranked probability density function. The primary goodness of fit criterion was the χ^2 or Chi-Square Test [212]. Where possible the Chi-Square test was seconded by the Kolmogorov-Smirnov (K-S) test statistic [212].

The Chi-Square test can be used with sample input data and any type of distribution function (discrete or continuous). A weakness of the Chi-Square test is that there are no clear guidelines for selecting intervals or 'bins'. To this end we experimented with a low and high resolution of sampling intervals and found heuristically the best interval by means of confirming the ranking of the statistical distribution achieved by the chi-square test together with the K-S test. This is because, the K-S test does not depend on the number of bins; such freedom makes the K-S test more powerful than the Chi-Square test, in cases where bin-dependencies can corrupt the accuracy of the ranking.

However, sole reliance to the K-S test was avoided, given that it does not detect tail discrepancies accurately. Where possible we further verified the distribution ranking with an additional Anderson-Darling test statistic (A-D) [213]. This test has the property of highlighting ranking differences between tails of different distributions while staying independent of the sampling interval, as opposed to the K-S test which focuses the energy of the statistic at the middle of the distribution [212].

To confirm the statistical correctness of the ranked distribution we further produced both a Probability (or P-P) and Quantile-Quantile (or Q-Q) plot. The P-P plot shows the p-value of the fitted distribution vs. the p-value of the fitted result. Where the fit is "good", the plot is nearly linear. A similar principle applies for the Q-Q plot but for quantile[17] values.

[17] By a quantile, we mean the fraction (or percent) of points below the given value. That is, the 0.3 (or 30%) quantile is the point at which 30% percent of the data fall below and 70% fall above that value.

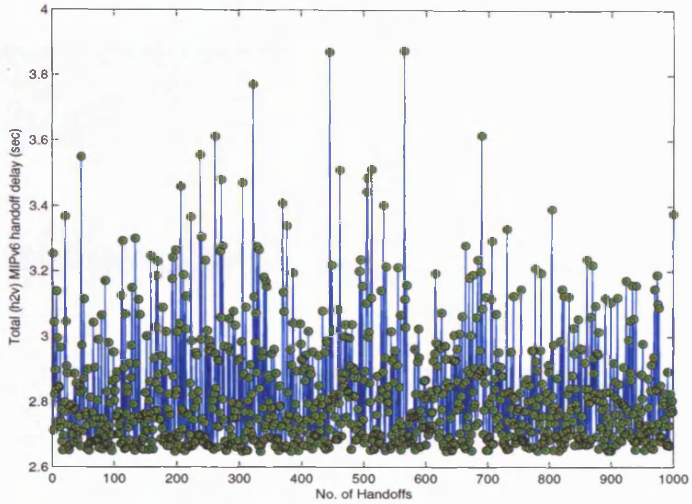


Figure 3.14: MIPv6 handoff delay experienced away from the home network

In the following sections we present the statistical distribution that scored the highest ranking among the set of tested distributions through the aforementioned goodness of fit criteria. We further provide the first statistical moments at a 95% confidence interval for two individual MIPv6 handoff cases during VoIP communications between MN and CN peers: (i) during a *h2v* MIPV6 handoff (away from home) or (ii) during a *v2h* MIPv6 handoff (returning home).

MIPv6 handoff performance away from Home network

Total MIPv6 handoff latency distribution From the delay measures obtained for 1000 handoffs we observed an exponential distribution of the total MIPv6 handoff delay derived from the sequence number of the received VoIP at the MN. 50% of the handoffs account for a delay of 2.751 sec.

The reported mean MIPv6 handoff delay of 2.8265sec maintains a confidence interval of 95% of the normalised handoff data as shown in the derived empirical density function of figure 3.16(a) with parameters $\beta = 0.17817$ and $shift = 2.64837$. The shift is applicable for the simple reason that the data measures exceed the range of the fitted distribution with a minimum handoff delay measure of 2.650sec.

Figure 3.16(b) shows the respective empirical cumulative distribution function (CDF) for the total handoff delay. We can observe that 95% of the handoffs experience MIPv6 handoff delays with an upper bound of 3.121 ms.

Figures 3.17(a) and 3.17(b) show the quality of the fitted distribution with respect to the measured data. We may observe that, with respect to upper quantile, the

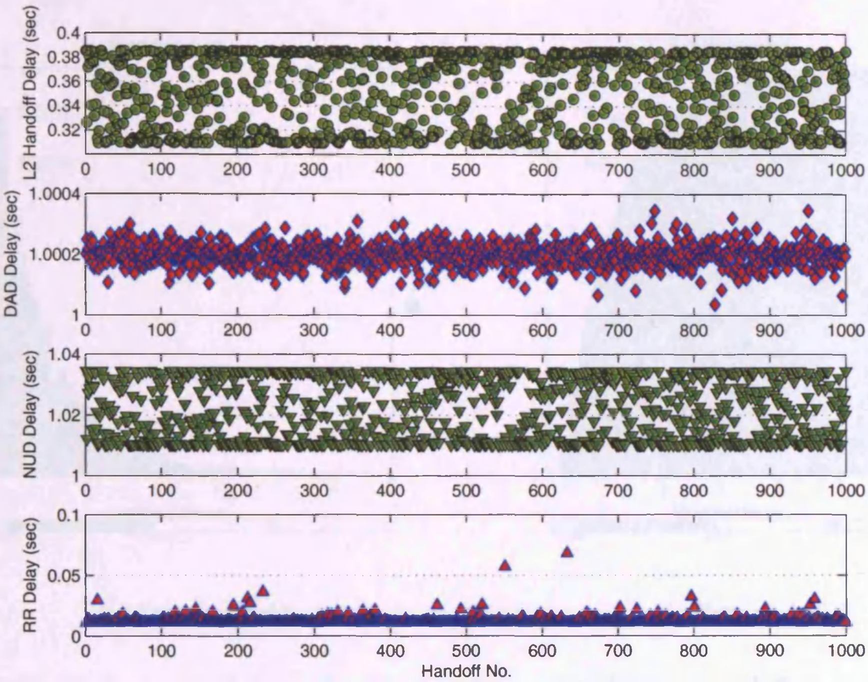


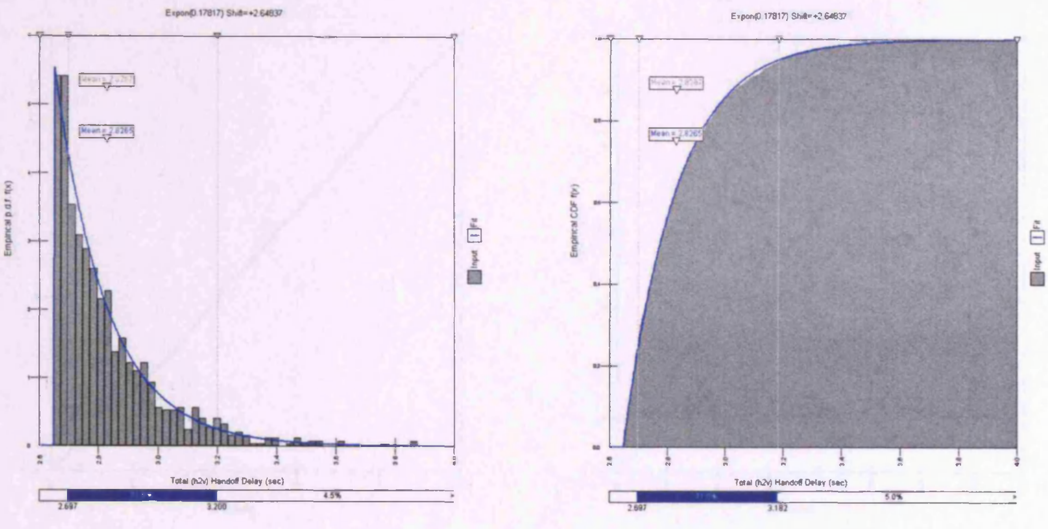
Figure 3.15: Identified MIPv6 handoff (2vh) delay components experienced away from the home network

measurement data maintain a smaller location value than the actual fitted distribution quantile. However, the probability plot validates that both input and fitted p-values are derived from a common distribution.

In subsequent sections, we describe the latency contribution by individual mechanisms collectively tracking the total delay of a MIPv6 handoff.

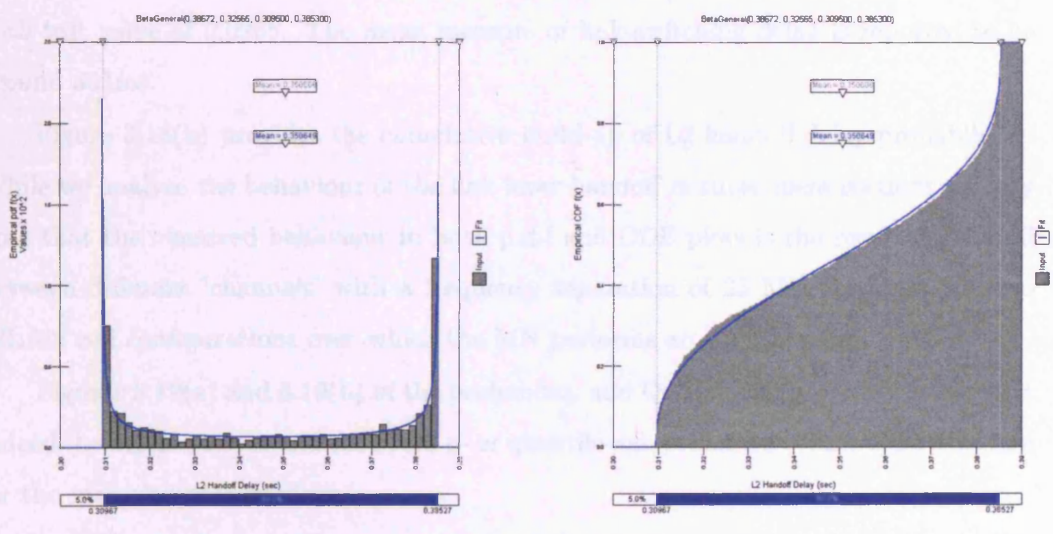
L2 handoff latency component With respect to the link latency component induced during link switching it can be seen that the link-switching component exhibits two peaks at the boundaries of observed link-layer delay measure; the peak reported at the lower bound of the empirical probability density function is found to be around 310ms while the upper bound peak found around 385ms.

Figure 3.18(a) present the plot of the empirical p.d.f derived together with the probability distribution indicated by the accompanying GoF set of test statistics conducted.



(a) Empirical p.d.f (b) Empirical CDF

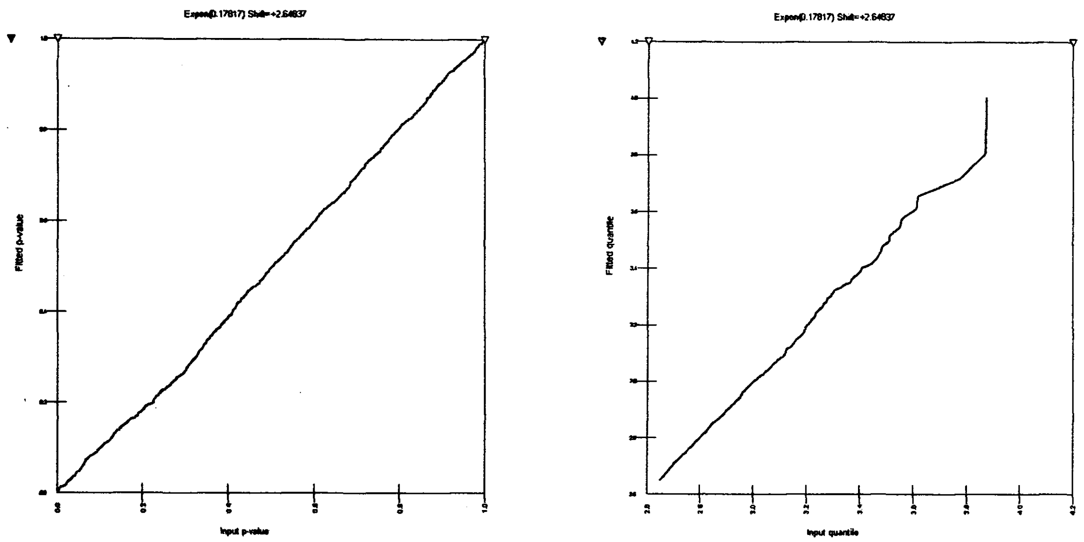
Figure 3.16: Total handoff delay distribution for a 1000 MIPv6 handoff transitions of the MN between visited networks



(a) Empirical p.d.f (b) Empirical CDF

Figure 3.18: L2-handoff delay distribution for a 1000 MIPv6 handoff transitions of the MN between visited networks

In particular, both Chi-Square and K-S test statistics agree that the particular set of measurement data follow a Beta distribution described by the parameters $\alpha 1 = 0.38672$ and $\beta = 0.32565$. With respect to GoF the Chi-Sq test value of 29.44 and p-value of



(a) Probability plot (b) Q-Q plot

Figure 3.17: P-P and Q-Q plots for total handoff delay according to the Chi-squared test result

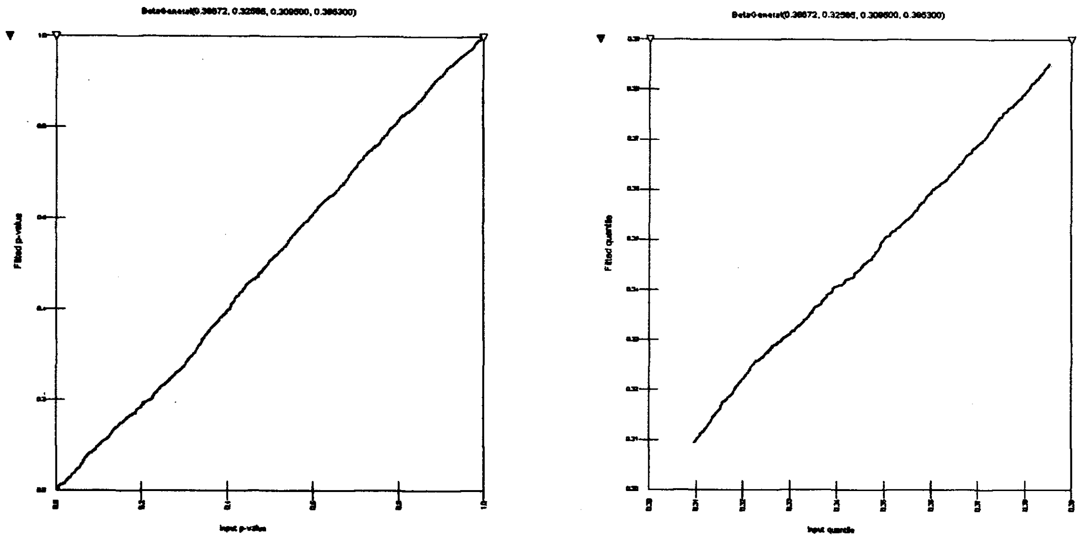
0.3904; the K-S test statistics confirms the same ranking in the particular distribution with test value of 0.0266. The mean measure of link-switching delay is reported to be around 351ms.

Figure 3.18(b) provides the cumulative build-up of L2 handoff delay probabilities. While we analyse the behaviour of the link-layer handoff in subsequent sections we may note that the observed behaviour in both p.d.f and CDF plots is the result of handoff between different 'channels' with a frequency separation of 25 Mhz amongst the two WLAN cell configurations over which the MN performs an IPv6 handoff.

Figures 3.19(a) and 3.19(b) of the probability and Q-Q plot respectively attest that indeed the input dataset and the fitted p- or quantile-values share a common distribution for the majority of the values.

Latency induced during Movement Detection In regards to movement detection during a MIPv6 handoff away from the home network, we have already identified three distinct cases of movement detection and their respective delay component, namely, nominal, sap-reactive and numb-reactive hangover delay.

In the case of the nominal hangover delay, the Chi Square test statistic (t-val =17.26, p-val=0.94) exhibits the highest ranking score for a log-logistic distribution with parameters α = 7.574 β = 0.064 and γ = -0.012. This is seconded also by



(a) Probability plot (b) Q-Q plot

Figure 3.19: P-P and Q-Q plots for L2 handoff latency component according to the Chi-squared test result

both K-S (t-val = 0.011) and A-D tests (t-val = 0.13) which are insensitive to bin size emphasising onto the tail of the distribution. The distribution exhibits a mean value of 53ms for both sample data and fitted values with a 90% confidence interval lying between 82 and 31ms as shown by the empirical p.d.f. of figure 3.21(a)

Figure 3.21(b) shows the respective cumulative distribution for the nominal hangover delay component. We may see that, the cumulative increase is nearly linear (albeit steep) between 22% and 75% of the handoffs accounted for this class of movement detection, between 41 and 59ms. So, a good 52% of this class of handoffs experience at most one lost periodic router advertisment while autoconfiguring a CoA address and prior to initiation of the DAD process.

Figures 3.22(a) and 3.22(b) show the probability and Q-Q plots of the derived nominal hangover distribution. The probability plot demonstrates a reasonably accurate fit, while the Q-Q plot reveals short tail at the right end of the distribution, given that the respective upper quantile (right end) is below the reference line $y = x$. This also indicates a few outliers at that tail, given these points do not fall on the reference line.

In the case of the sap-reactive hangover delay, tracked by the loss of a router solicitation, the Chi Square (p-value=0.999) test attests that this hangover day is Pareto-distributed with parameters $q = 12.665$ and $\alpha = 0.26660$. This is seconded also by and

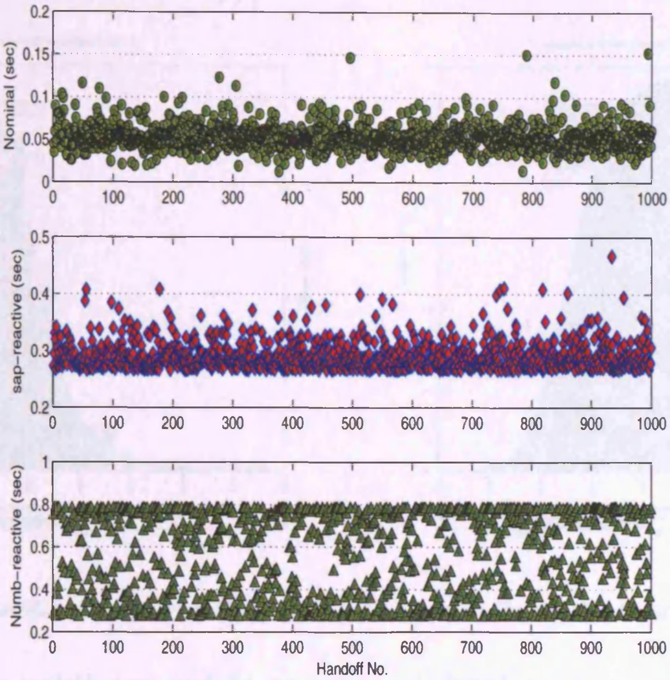


Figure 3.20: Numb-reactive, Sap-reactive and Normal types of hangover delay experienced by the MN away from the home network

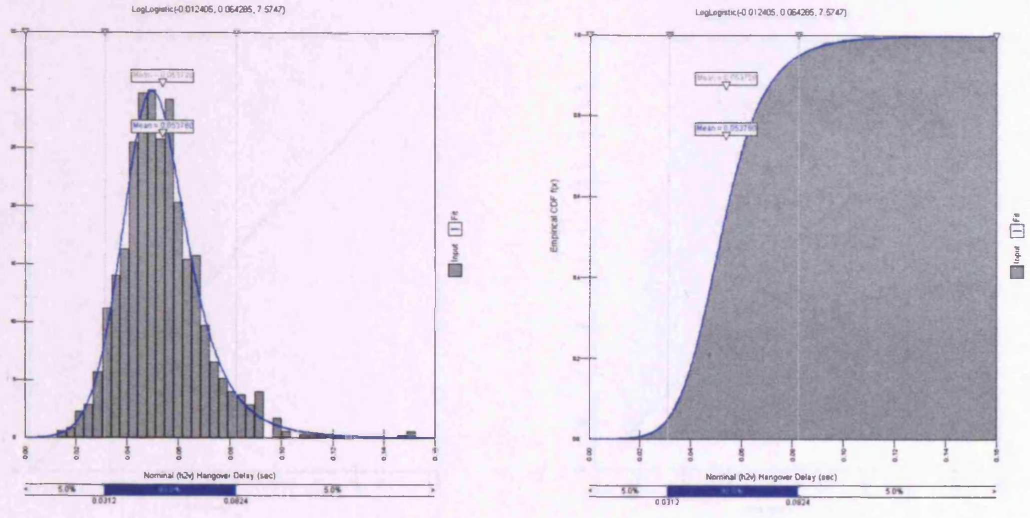
additional K-S (t-val = 0.011) test with respect to the tail of the distribution. The Pareto-distributed dataset has a mean value of 289ms for both sample data and fitted values with a 90% confidence interval lying between 268 and 338ms as shown by the empirical p.d.f. of figure 3.23(a)

Figure 3.23(b) shows the respective cumulative distribution for the nominal hangover delay component. We may see that, about 50% of these MIPv6 handoffs experience a delay of up to 280ms experiencing a nearly-linear increase as the number of handoffs increases. The mean delay represents about 75.6% of the number of handoffs experiencing this type of hangover delay as the MNs attempt to autoconfiguring its IPv6 connectivity state and prior to initiation of the DAD process.

Figures 3.24(a) and 3.24(b) show the probability and Q-Q plots of the derived Pareto distribution experienced during a sap-reactive movement detection. Both probability and Q-Q plots demonstrate a reasonably accurate fit, with a few outliers.

Analysing statistically the case of a numb-reactive hangover delay, we find that the interval that the next router solicitation is sent out varies between 270 and 585ms according to a beta distribution, as shown in figure 3.25(a) with location[18] $\alpha = 0.336$

[18]the location parameter typically shifts to the right/left the fitted curve relative to the reference probability distribution



(a) Probability Density Function (b) Cumulative Distribution Function

Figure 3.21: Nominal Hangover delay component induced during movement detection by MIPv6 handoff process

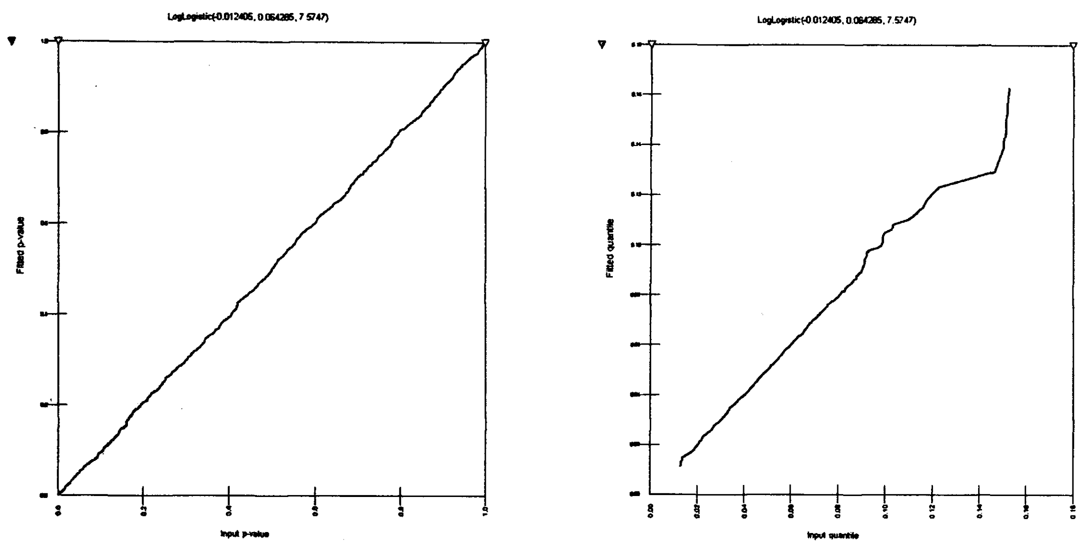
and scale[19] $\beta = 0.439$.

Figure 3.25(b) shows the build-up of these probabilities $(P(k \leq x))$ in the respective cumulative distribution function derived.

The mean delay value imposed by this handoff component function is around 408ms. Figures 3.26(a) and 3.26(b) demonstrate that the fit is not particularly good, i.e. does not describe accurately the behaviour of the sampled dataset with respect to the particular delay component. This is despite the fact that the GoF tests indicated a beta distribution as the prime candidate for describing the behaviour of this particular delay component.

DAD For the MN roaming away from its home network, the delay incurred by DAD process is tracked by the RETRANS_TIME as per the stateless address auto-configuration standard [110]. Our measurements confirm that, this is indeed the case: 1000ms are consumed consistently for all 1000 handoffs, as soon as the neighbour solicitation is sent out by the MN to announce the allocation of the particular tentative CoA to itself. A negligible standard deviation of 2ms is observed which may be accounted to interrupt-driven delays incurred by other IPv6 messaging such as the periodic receipt of Router

[19]the effect of a scale parameter < 1 is to compress the p.d.f., i.e. compress the density of the probabilities to a smaller range than the reference distribution. If the scale is > 1 the effect is stretching the density of probabilities to a larger range



(a) Probability plot (b) Q-Q plot

Figure 3.22: P-P and Q-Q plots for Nominal Hangover delay component induced during movement detection by MIPv6 handoff process

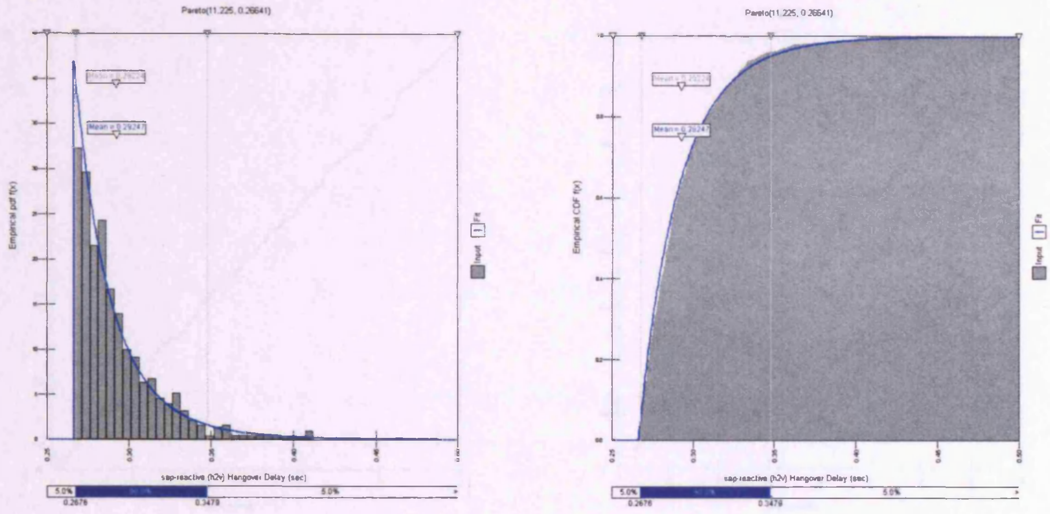
Adverts.

Neighbour reachability handoff latency component Figure D.2(a) present the empirical p.d.f derived for the latency component pertaining to the neighbour reachability.

To arrive at the observed distribution we first conducted the Ch-Square statistic to evaluate the GoF on the set of parametric distribution families for a range of 10 equiprobable intervals (bins). The test (t-value=34.6, p-value=0.18) showed the beta distribution to achieve the highest GoF ranking amongst the probability distributions tested[20]. This probability distribution is described by a location, $\alpha 1 = 0.281$ and scale, $\beta = 0.323$. We repeated the same GoF test for a higher sampling resolution (30 bins) observing no significant difference in the distribution ranking; this is also confirmed by the K-S test with a t-value of 0.039.

The distribution for this delay component attains a mean of 1.021sec and a std. dev of 9.6ms. Both probability (fig. D.3(a)) and Q-Q plots (fig. D.3(b)) indicate that both fitted and actual dataset values draw from the same distribution. Figure D.2(b) shows the build-up of these probability densities for $P(k \leq x)$ in the respective CDF derived.

[20] All known distributions were tested by means of both Stata and Matlab statistical analysis tools



(a) Probability Density Function (b) Cumulative Distribution Function

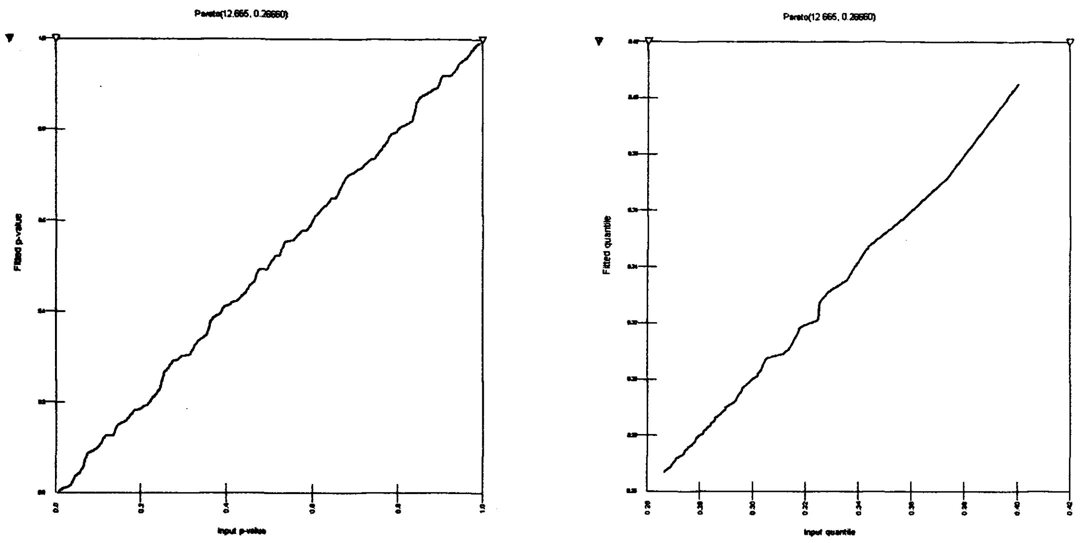
Figure 3.23: sap-reactive hangover delay component induced by the MIPv6 handoff process, by loss of first router solicitation and insensitivity to RtAdv messages during movement detection

We may note that neighbour reachability *coincides* with the CoA registration period T_r imposed by the update of bindings with the HA. It can be seen from equation (D.8) that T_r is primarily tracked by the RTT between MN-HA, since BU processing delay on today's fast processors range between 1-2ms, as attested by subsequent sections.

A number of studies have conferred on Internet dynamics and RTT variability [214], with the latter being realistically approximated as a shifted Gamma distribution [215, 216].

Recent studies in [217, 218] with focus on VoIP traffic over fast paths in North America or transatlantic links, confirm this type of distribution with reported a mean *end-to-end* (one-way) delay in the order of 75-80ms. Such delay is not uncommon over a number of US provisioning domains and occur as a result of non-optimal routes followed between the communicating parties in an effort to balance congestion [219]. This translates to a worst-case RTT of 150-160ms, which, in turn, brings the mean T_r registration delay to a worst case of 152-162ms. This is clearly significantly smaller that the neighbour reachability delay reported in our measurements; thus T_r becomes overshadowed by magnitude of T_{NUD}.

What is important, however, is that *reliance of the IPv6 handoff process on the*



(a) Probability plot (b) Q-Q plot

Figure 3.24: P-P and Q-Q plots for hangover delay component induced by the MIPv6 handoff process, by loss of first router solicitation and insensitivity to RtAdv messages during movement detection

completion of either neighbour reachability or binding registration with the HA, is guaranteed to impact significantly the seamlessness principle with respect to both interactivity and intelligibility of a VoIP conversation for the MN on the move.

Binding Updates From the perspective of Binding Updates towards the CN peer we observe a mean delay of 4.9ms for the HoT and 7.6ms for the CoT message sequences with std.dev of 1.2ms and 1.5ms respectively. Given that the RTT on the wireless link at 1 Mbps has an average measure of 4.2ms, we can see that the HoT sequence results about 0.7ms total processing time; for the CoT sequence, given that the RTT on the fast Ethernet interface is into the sub-millisecond range, the total processing time amounts to 3.3ms. In both, cases processing time compared to the actually RTT experienced in the Internet can be seen to be negligible.

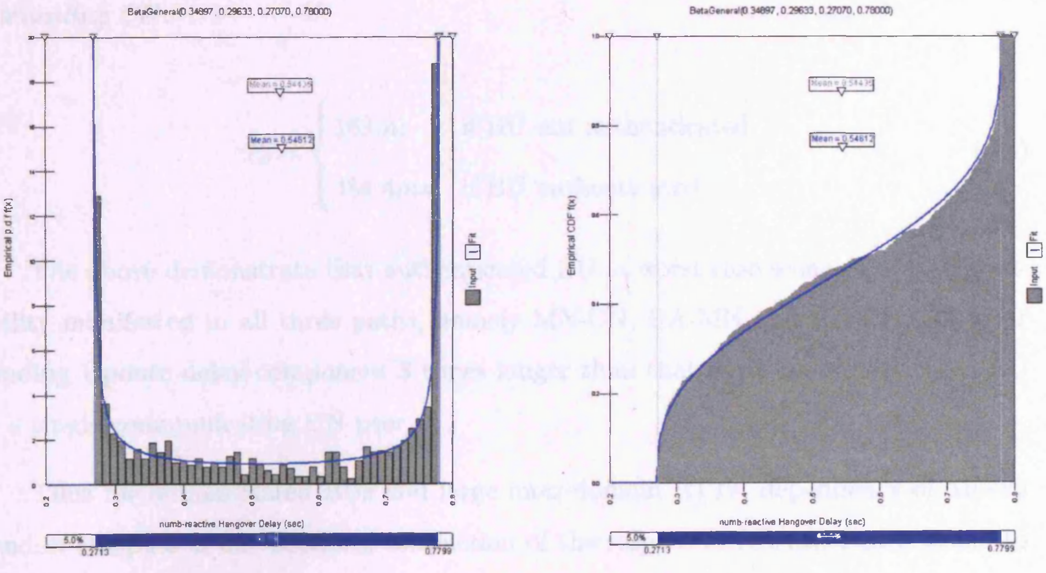
The measure of total handoff delay described so far does not incorporate any realistic measure of end-to-end delay as experienced in the Internet today, given that the RTT between the MN and its peers during our measurement tests is negligible ($4.2ms$). This is because this measurements set attempts to account in detail for the individual components of MIPv6 handoff delay without influences from network externalities such as RTT variability.

From this perspective the binding update process of the MIPv6 total handoff delay would increase by an additional RTT. A coarse approximation of 1 RTT would be 2 times the end-to-end delay experienced between the two peers. Such approximation may suffice for worst case performance purposes but it does not necessarily reflect the accurate measure of delay for that RTT. This is because during a transmission the forward and return path between two peers are not necessarily the same. In fact, it is possible to route either path of VoIP traffic between peers (either forward or return) through routes which - while not optimal - actually incur smaller end-to-end delays [220].

Considering the worst case end-to-end delay scenario discussed in the previous section, as the expected upper delay bound on all accounted RTT between signalling interactions, and by using equations (D.8) and components (D.9) and (D.10), we have:

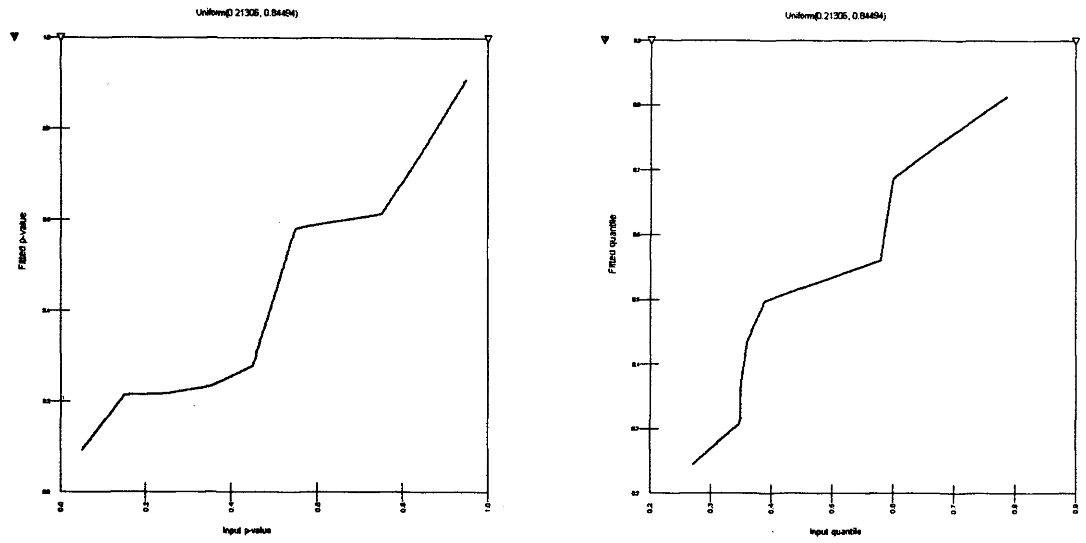
$$T_o = \begin{cases} 163ms & \text{if BU not authenticated,} \\ T_{RR} + 163ms & \text{if BU authenticated} \end{cases} \tag{3.2}$$

$$T_{RR} = \begin{cases} 321.4ms & \text{if } RTT_{RR} > RTT_{MN-CN}, \\ 163.3ms & \text{if } RTT_{RR} < RTT_{MN-CN} \end{cases} \tag{3.3}$$



(a) Probability Density Function (b) Cumulative Distribution Function

Figure 3.25: numb-reactive hangover delay during the respective type of movement detection in MIPv6 handoffs away from the home network



(a) Probability plot (b) Q-Q plot

Figure 3.26: P-P and Q-Q plots for numb-reactive hangover delay during a MIPv6 handoff

Given that in the worst case of RTT delay the $RTT_{RR} = 320ms$ is greater than $RTT_{MN-CN} = 160ms$, then the delay incurred by return routability message sequence becomes $T_{RR} = 321.4$. This results effectively the following T_o delay *for a single communicating CN*:

$$T_o = \begin{cases} 163ms & \text{if BU not authenticated,} \\ 484.4ms & \text{if BU authenticated} \end{cases} \qquad (3.4)$$

The above demonstrate that authenticated BU at worst case scenario of RTT variability manifested in all three paths, namely MN-CN, HA-MN and HA-CN will incur Binding Update delay component 3 times longer than that of an unauthenticated BU to a single communicating CN peer.

Thus for authenticated BUs and large inter-domain RTTs, dependency of MIPv6 handoff completion on successful completion of the respective RR function is found to impact significantly the measure of experienced MIPv6 handoff delay, in addition to the delay components presented thus far. Hence, *BU authentication can degrade significantly the quality of the VoIP session by introducing inflated one-way delays between the MN and CN during MIPv6 handoff completion.*

MIPv6 handoff delay summary away from the home network

A summary of first statistical moments for a MIPv6 handoff away from the home network is shown in table 3.2. The table provides the handoff delay components identified by experimentation.

The table distinguishes between the nominal hangover delay period accounted in 17.8% of the total number of handoffs measured and sap-reactive (sap-R) as well as numb-reactive (numb-R) hangover delay accounted for, in movement detection during a MIPv6 handoff. The two latter cases account for insensitivity to multicast periodic RtAdv messages *with* (30.98% of MIPv6 h2v handoffs) or *without* (51.22% of MIPv6 h2v handoffs) potential loss of the router solicitation incurred during the L2-handoff period.

It thus, becomes clear that the case of h2v MIPv6 handoffs introduces significantly higher MIPv6 handoff delay in contrast to the v2h MIPv6 handoff, elaborated in the following section. In either case, both types of handoff incur significant IPv6 handoff latencies sufficient to disrupt VoIP communications.

MIPv6 Handoff	min	max	50%	95%	Mean	Std Dev	Variance
delay component	(sec)						
T_{l2} (all cases)	0.308	0.536	0.407	0.501	0.423	0.086	0.0024
T_{ho} (nominal)	0.012	0.152	0.051	0.082	0.053	0.016	2.6959E-05
T_{ho} (sap-R)	0.266	0.678	0.281	0.337	0.289	0.024	6.2027E-5
T_{ho} (numb-R)	0.270	0.587	0.375	0.585	0.404	0.118	0.013
T_{DAD} (all cases)	1.000	1.000	1.000	1.000	1.000	0.000	0.000001
T_{NUD} (all cases)	1.010	1.034	1.019	1.034	1.020	0.009	8.3479E-05
T_{hoff} (nominal)	2.353	2.594	2.436	2.568	2.463	0.046	0.00136
T_{hoff} (sap-R)	2.592	3.289	2.654	2.802	2.715	0.147	0.021
T_{hoff} (numb-R)	2.65	3.549	2.751	3.121	2.803	0.161	0.025

Table 3.2: Statistical moments for MIPv6 (h2v) handoff delay away from the home network and its components at 90% confidence interval

From this class of handoffs it interesting to note that, *even if the DAD process is dropped in the light of very low probability of configuration of a duplicate IPV6 address, the delay induced by neighbour reachability would persist until the cache entry has reached the REACHABLE state.* Such behaviour is consistent in all 1000 handoff conducted during our experiments.

Thus the host of delay components during an IPv6 handoff is dominated as seen above not only from the DAD process but also from the delay incurred by the current form of neighbour reachability. The delay distribution of the MIPv6 handoff delay component owed to neighbour reachability signalling is presented in Annex D.4.

MIPv6 handoff performance returning to Home network

Figure 3.27 shows the respective total delay trace for all 1000 handoffs recorded during this series of experiments. Table 3.3 presents the respective first statistical moments at a 90% confidence interval.

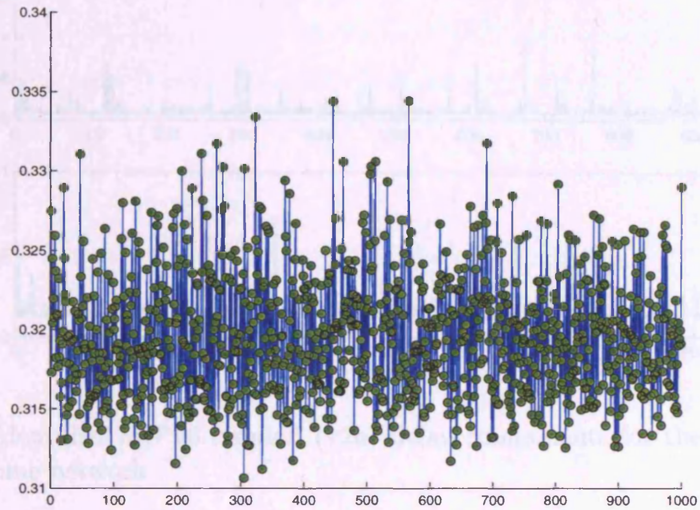


Figure 3.27: Total MIPv6 handoff (v2h) delay for the MN returning back to the home network

Figure 3.28 shows the individual delay components identified as part of the total handoff delay presented in figure 3.27

In Annex D.5 we present the respective fitted distribution as well as probability and quantile plots, for the measure of MIPv6 handoff delay of a v2h MIPv6 handoff. It may be seen that upon return to the home network most delay components are significantly smaller in contrast to the ones of an (h2v) MIPv6 handoff away from the home network.

MIPv6 Handoff	min	max	50%	95%	Mean	Std Dev	Variance
delay component	(sec)						
T_{l2}	0.241	0.380	0.322	0.380	0.376	0.053	0.0028
T_{DAD}	0.007	0.063	0.008	0.021	0.012	0.007	5.71E-05
T_{NUD}	0.012	0.043	0.012	0.022	0.016	0.006	4.450E-05
T_{hoff}	0.395	0.422	0.406	0.414	0.406	0.004	1.628E-05

Table 3.3: Statistical moments for MIPv6 (v2h) handoff delay on return to the home network and its components at 90% confidence interval

From the captured v2h and h2v MIPv6 handoff traces we have further analysed the performance of the supporting Neighbour Discovery protocol. For reasons of space

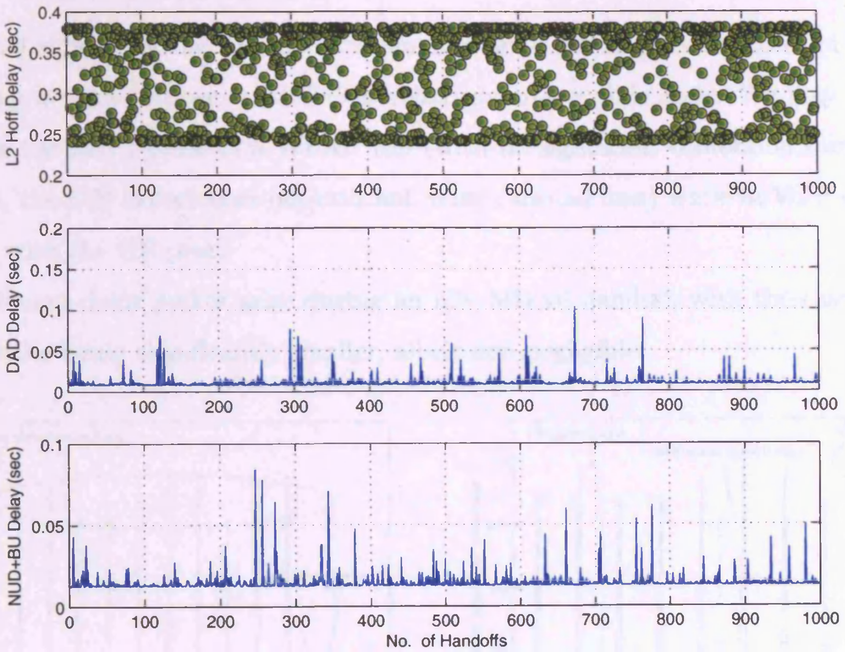


Figure 3.28: Identified MIPv6 handoff (v2h) delay components for the MN returning back to the home network

conservation and brevity this is presented in Annex D.5.1.

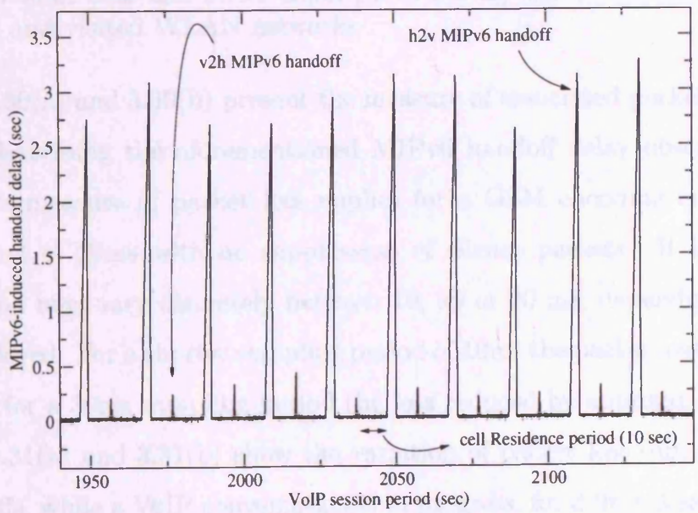
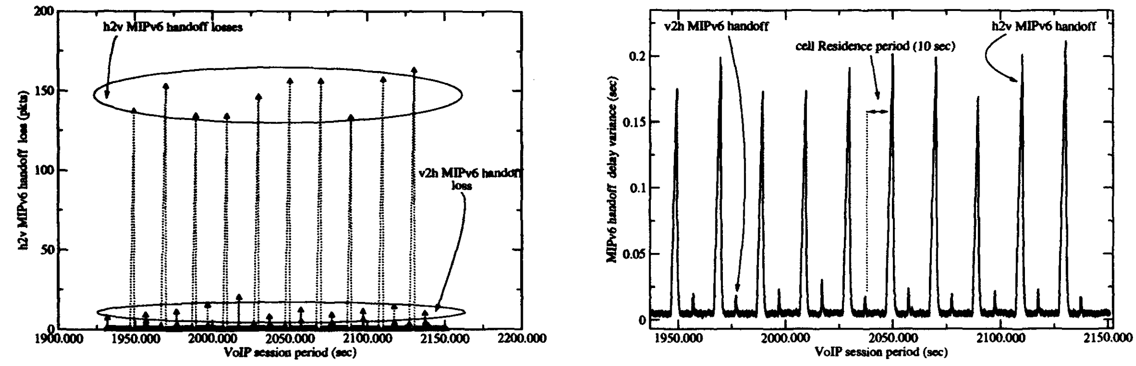


Figure 3.29: Overall MIPv6-enabled handoff while in VoIP communications for the MN on the move.

3.8 Packet Loss and Jitter

Figure 3.29 shows the effect of MIPv6 handoffs in terms of delay induced on a VoIP flow during a voice call over a MIPv6-enabled roaming mobile node. We may observe that while the MN resides in a WLAN cell (with no significant contention over its air interface), the MN experiences insignificant delay (around 5ms) while in VoIP communications with the CN peer.

Significant delay spikes arise during an h2v MIPv6 handoff, with the case of v2h handoff delay being significantly smaller, albeit non-negligible.



(a) Loss

(b) Jitter

Figure 3.30: Overall Loss and Jitter experienced during MIPv6 enabled IPv6 roaming between home and visited WLAN networks

Figures 3.30(a) and 3.30(b) present the measure of associated packet loss and jitter respectively, describing the aforementioned MIPv6 handoff delay observed. We note that the above measure of packet loss applies for a GSM encoding and a respective sampling period of 20ms with no suppression of silence packets. It is possible that sampling period may vary discretely between 10, 20 or 30 ms, depending on the VoIP encoding employed. For a shorter sampling period of 10ms the packet loss approximately doubles while for a 30ms sampling period the loss reduced by approximately 33%.

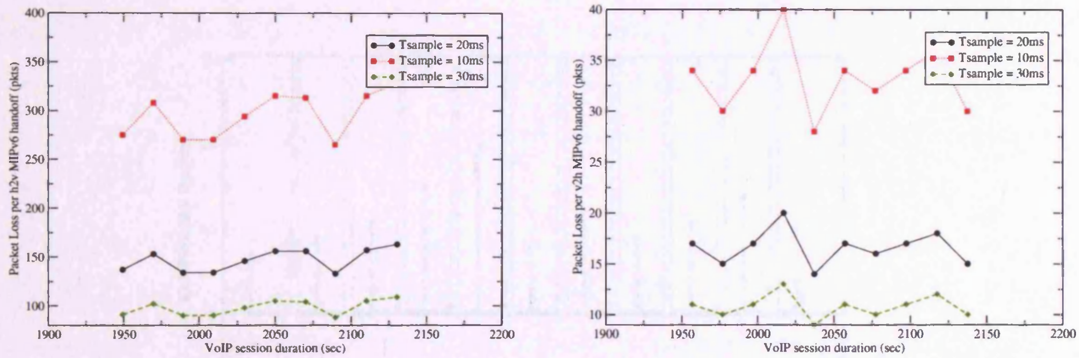
Figures 3.31(a) and 3.31(b) show the variation of packet loss during h2v and v2h MIPv6 handoffs, while a VoIP conversation is in progress, for different sampling periods of VoIP encodings.

Further investigation of MIPv6 performance under different encodings is considered out of scope for the purposes of this thesis and as such is considered as future directions of research.

It is interesting to note that during the MIPv6 handoff (v2h) on return back to the home network the VoIP flow session indicates more than one cluster of lost packets. The same is not arising in the case of the h2v handoff. We consider the most probable cause for such behaviour, the particular implementation of the IEEE 802.11b AP, given that the AP on the home network features a more recent firmware OS.

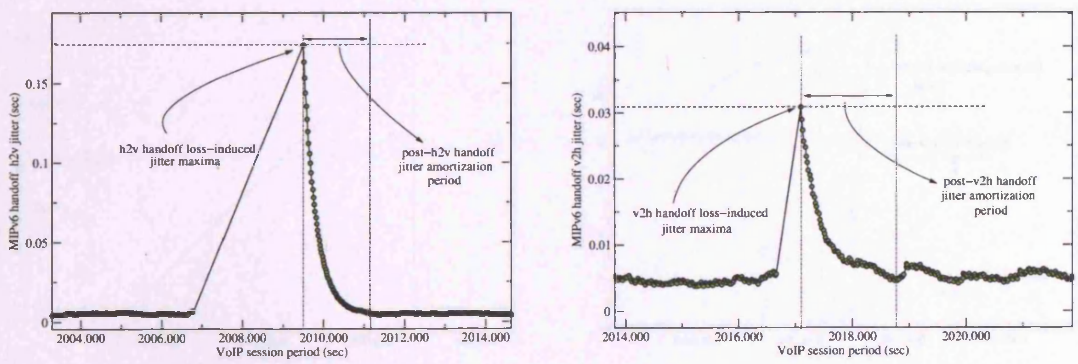
Figures 3.34(a) and 3.34(b) show the loss-run signature of h2v and v2h MIPv6 handoffs.

For this particular firmware OS version, on completion of the scan phase, the MN appears to return back to its original channel before it effects the L2-handoff decision and subsequent association with the new AP. The IEEE 802.11 specification [17] does not specify whether the wireless station must stay off the existing channel under association, hence vendor implementations are free to optimise the particular function as they see



(a) h2v Loss (b) v2h Loss

Figure 3.31: Packet loss for h2v and v2h MIPv6 handoff cases for the duration of a single VoIP session for different voice sample periods



(a) h2v Jitter (b) v2h Jitter

Figure 3.32: Jitter pattern for h2v and v2h MIPv6 handoffs as well as required jitter amortisation period before VoIP session returns to normal jitter levels.

fit.

The above is confirmed by merely exchanging the AP devices between the two neighbouring cells. We repeat the experiment, observing the loss pattern on the visited network under this setup. Indeed, clustering of packet losses and an identical L2-handoff loss-run signature arises also in the case of the h2v handoff.

From the above we may deduce however, that heterogeneity between AP implementations, while not offending standard specifications can result into different patterns that the L2 handoff may be effected. This does not, in effect, alter the total L2-handoff delay, but *may* induce shorter loss runs in the VoIP communication stream towards the MN. If the size of loss runs incurred, during the L2-handoff process, can be sustained to ≤ 2 packets then PLC techniques may potentially sustain the VoIP quality by replenishing the lost packets artificially.

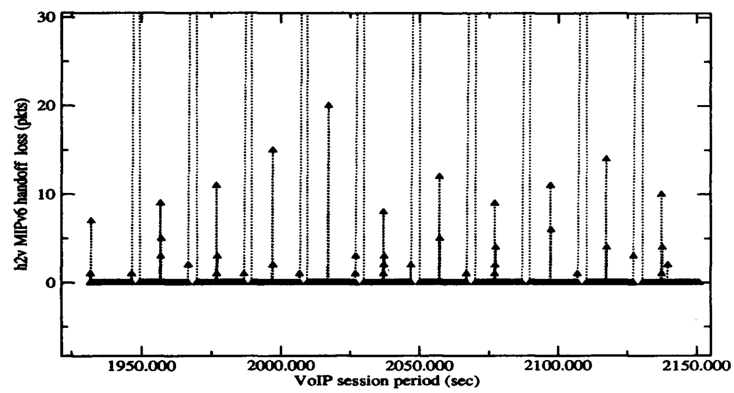
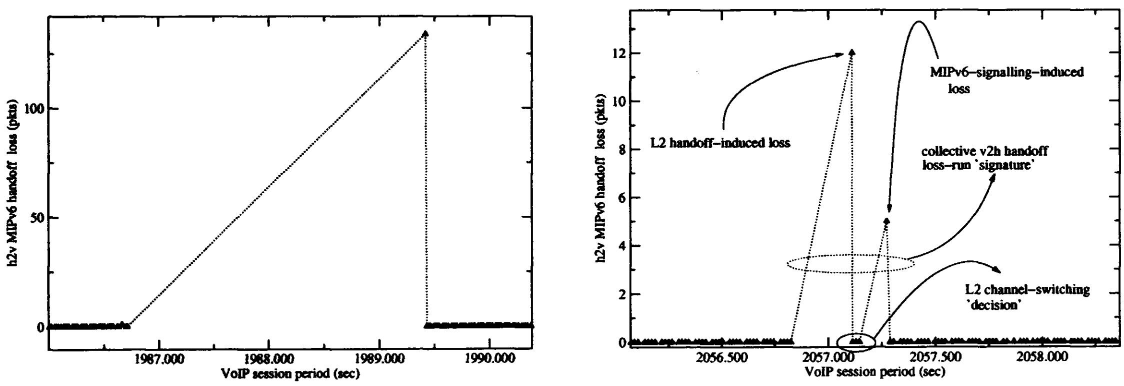


Figure 3.33: Loss Run clusters observed during MIPv6 handoffs returning home. This may be associated with WLAN AP optimisations



(a) h2v Loss Run Signature

(b) v2h Loss Run Signature

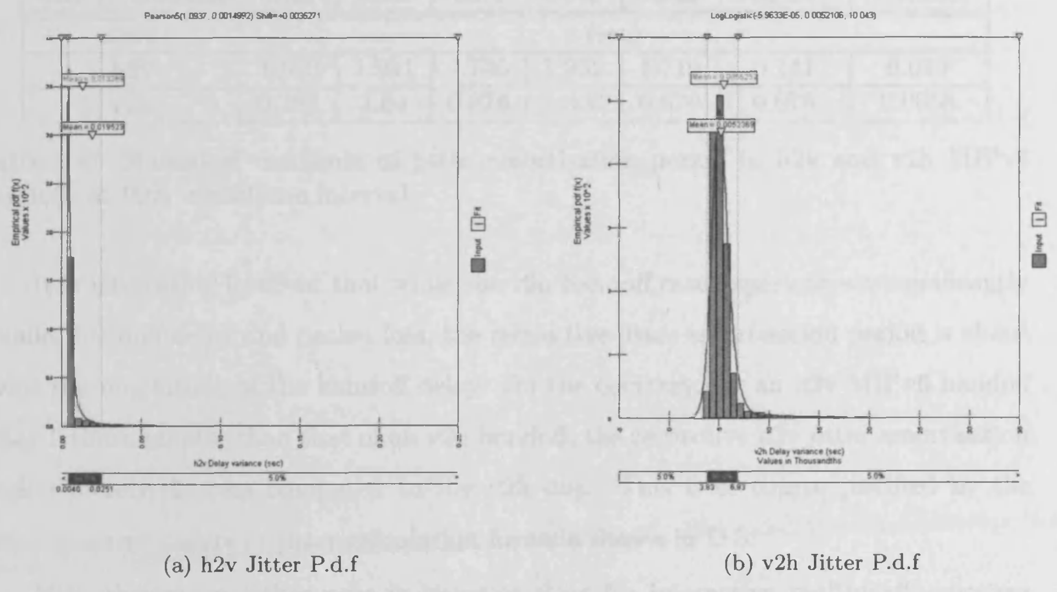
Figure 3.34: Identified Loss Run signatures characterising the handoff process for h2v and v2h MIPv6 handoffs

Figures 3.35(a) and 3.35(b) show the distribution of delay variance while in VoIP communications, during h2v and v2h MIPv6 handoffs respectively. We may observe that the h2v handoff case have an average measure of jitter 4 times higher than the average measure of delay variance in the v2h case of handoff. Such difference is clearly proportional to the magnitude of total handoff delay between h2v and v2h handoffs respectively, given that handoffs are effected with the same frequency in both h2v and v2h roaming cases.

We may observe also that after the completion of a MIPv6 handoff the jitter peak experienced requires a minimum period, termed as *jitter amortisation period*, before the jitter average returns back to its nominal levels.

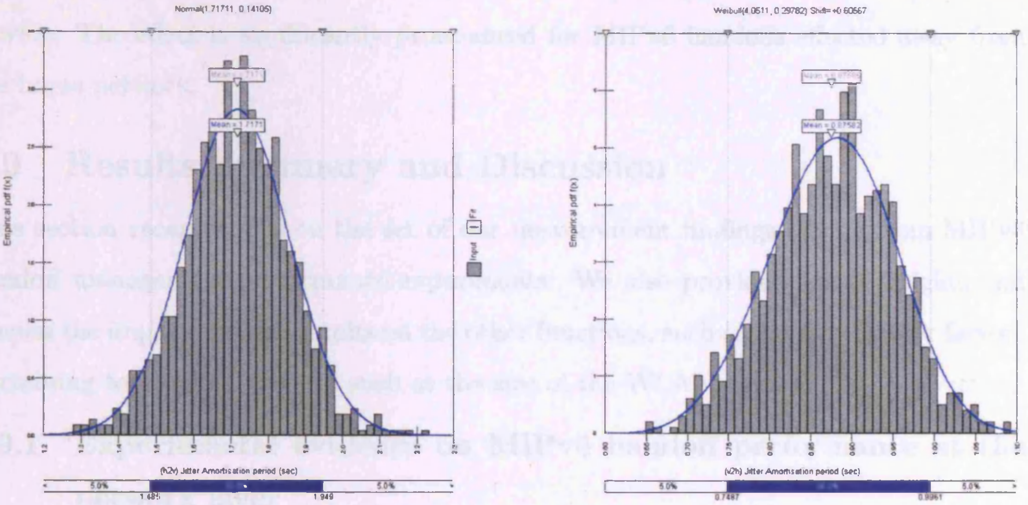
To identify the average jitter amortisation period for each of the two MIPv6 handoff cases, we first derive the mean delay variance for a 90% confidence interval; we then subtract from the instantaneous jitter measure observed through the series of handoff measurements. We then plot the delta in two cases (i) for the lowest mean jitter observed in the case of the v2h handoff and (ii) for the highest mean jitter observed in the case of the h2v handoff.

Figures 3.36(a) and 3.36(b) show the respective probability density of the Jitter amortisation period essential during an h2v and v2h MIPv6 handoff respectively. It can be seen that for large handoff delays and and their associated loss-runs (h2v handoff),



(a) h2v Jitter P.d.f (b) v2h Jitter P.d.f

Figure 3.35: Jitter distribution for h2v and v2h MIPv6 handoff cases.



(a) h2v Jitter P.d.f (b) v2h Jitter P.d.f

Figure 3.36: Jitter amortisation period distribution for h2v and v2h MIPv6 handoff cases.

the respective jitter amortisation period increases; for smaller handoff delays (mean of 406ms in the v2h handoff case) and resulting loss-runs (v2h handoff) the jitter amortisation period decreases (mean of 877ms). Table 3.4 provides the summary statistics describing the jitter amortisation period in MIPv6 handoffs away from and returning to the home network.

MIPv6 Handoff	min	max	50%	95%	Mean	Std Dev	Variance
case	(sec)						
h2v	1.503	1.951	1.735	1.952	1.719	0.141	0.019
v2h	0.784	1.04	0.876	1.003	0.876	0.076	0.0058

Table 3.4: Statistical moments of jitter amortisation period in h2v and v2h MIPv6 handoffs at 90% confidence interval

It is interesting however that while the v2h handoff case experiences a significantly smaller handoff delay and packet loss, the respective jitter amortisation period is about twice the magnitude of the handoff delay. On the contrary, for an h2v MIPv6 handoff delay 5 times greater than that of an v2h handoff, the respective h2v jitter amortisation period merely doubles compared to the v2h one. This is of course justified by the autoregressive nature of jitter calculation formula shown in D.3.

Nevertheless, in either case it becomes clear for interactive multimedia services MIPv6 handoffs, incur significant jitter that can impact negatively lip synchronisation between coupled audio/video streams [25, 221] comprising the interactive multimedia service. The effect is significantly pronounced for MIPv6 handoffs effected away from the home network.

3.9 Results Summary and Discussion

This section recapitulates on the set of our measurement findings arising from MIPv6 handoff management performance experiments. We also provide critical insights and discuss the impact of these results on the other functions, such as AAA or QoS or factors pertaining to mobility pattern such as the size of the WLAN domain.

3.9.1 Experimental evidence on MIPv6 handoff performance at the network layer

From our measurements regarding the collective behaviour of Mobile IPv6 management specification, we may deduce that:

- Duplicate address detection incurs a significant amount of delay consistently for all handoffs measured.
- While the router advertisement interval is shortened to 50ms, we find that for a significant amount of the handoffs measured, the handoff away from the home network does not exploit it during movement detection. It relies essentially on the prompt response to a router solicitation with a unicast router advertisement.
- the timing of a router solicitation during a MIPv6 handoff is inappropriate, in a way that it impacts significantly the movement detection process (inducing the sap-reactive hangover delay component) and thus impedes the prompt completion of the MIPv6 handoff process.
- the rate of router solicitations during movement detection is such that impedes an expedient movement detection. As a result the MIPv6 movement detection delay impacts in turn the handoff process, incurring significant delay.
- response to router solicitations, excluding the one that originated during an L2-handoff in progress consume typically 50-60ms, before a unicast router advertisement is sent to the MN.
- An authenticated binding update to a single corresponding node incurs significant delay that is guaranteed to increase the delay of MIPv6 handoff by about 484ms,

in the worst case RTT scenario between the MN-HA, MN-CN and HA-CN host pairs.

- The delay of a MIPv6 handoff for the MN return back to the home network (v2h MIPv6 handoff) is significantly smaller than the respective delay incurred during a handoff to a visited network (h2v MIPv6 handoff).

- The significantly smaller measure of MIPv6 handoff delay on an v2h MIPv6 handoff is accounted by the existence of the HA entity co-located on the same host as the Access Routing function. This implies, that the MIPv6 protocol effectively requires the existence of the HA function at the router devices to allow for low MIPv6 handoff delays incurred by the Mobility-enabled IPv6 layer[21].

- The delay imposed by the neighbour unreachability detection process is significant, so as to impact receipt of the Binding acknowledgement from the HA as well as the subsequent update of bindings (with or without authentication) at the communicating peers.

- Post-movement-detection hangover delay contributes a significant delay component to the total MIPv6 handoff latency. While this may be possibly accounted by inefficiencies in the implementation of the IPv6 specification (i.e. neighbour discovery), it attests that for different mobile devices, potential inefficiencies (induced heterogeneity) in IPv6 neighbour discovery has a cascading effect on the total MIPv6 handoff delay component.

- From the perspective of packet loss for a packetisation rate of 20ms and no suppression of silence packets, an h2v handoff incurs loss runs above between 150-154 packets. While voice activity detection (VAD) can reduce the amount of packet loss within the same period of handoff delay, the amount of reduction is dependent on the degree of activity of the voice conversation between the participants or the composition of a voice conversation. For experimental purposes VAD-enabled VoIP flows are left as future work for the purposes of this investigation.

- MIPv6 incurs significant jitter during a MIPv6 handoff. The measure of jitter induced is accounted by actual packet loss during the MIPv6 handoff, and is capable to offset significantly lip synchronisation in potential interactive multimedia that comprise of both audio and video. The measure of the MIPv6 handoff and imposes

[21]that is excluding the L2 handoff delay which is owed to delays incurred by the wireless technology.

a minimum jitter amortisation period before which any subsequent handoff will simply worsen the amount of jitter experienced as well as the loss of lip synchronisation. We remind that these packets cannot be recovered by either increases play-out delay at the receiver or by packet loss concealment techniques

With respect to delay incurred at the link layer of IEEE802.11, as the wireless technology of choice during this experimental measurements, we may first acknowledge that this delay component is independent of the MIPv6 handoff delay observed at the IPv6 layer. At its current form we may conclude that it affects the prompt completion of a MIPv6 handoff in the following ways:

- The MAC layer of the 802.11b WLAN specification incurs significant delay which by itself can impede any guarantees of real-time delivery for VoIP packet flows between the MN and its peers.
- The dominant delay component in an L2-handoff is incurred by the AP-discovery phase, as the MN scan reactively the channels triggered by a low SNR threshold.
- During an L2-handoff *all* channels are scanned before an AP is selected. While the algorithm of AP-discovery and in particular the ordering in which channels may be scanned is not mandated by the protocol, it is clear that the scanning process remains agnostic of the candidate APs surrounding the MN, unless explicitly scanned *in reaction* to signal loss below a certain SNR threshold.
- For an increasing number of associated MNs under the same AP, any reliance of the MIPv6 to small measures of router advertisement interval proves detrimental to the completion of the movement detection process and as a result the completion of the MIPv6 handoff. This is because the size of the advertisement interval is offset by a significant amount of frame delay which effectively results into a prolonged movement detection before the first router advertisement is received.

It may be seen that even by ignoring the L2-handoff delay component incurred by the MAC layer of 802.11b, the delay incurred by the MIPv6 layer alone is sufficient to place a VoIP flow below any guarantees of interactive real-time delivery of IPv6 traffic.

The above provide overwhelming evidence that the MIPv6 protocol standard incurs significant delay at the IPv6 layer, such that it cannot preserve the seamlessness principle so as to sustain real-time guarantees in the delivery of interactive multimedia services to wireless IPv6 mobile devices.

It is intuitive that any discussion of additional IP connectivity context, such as AAA or QoS is bound to impact the completion of an IPv6 handoff[22] by an additional delay component which is bound by the worst RTT experienced between the communicating entities effecting AAA or QoS control signalling if setup delay is assumed to be negligible.

While in both v2h and h2v handoff cases there exists a significant L2 handoff delay components of about 420ms which is beyond the control of MIPv6 mobility management, the MIPv6 mechanism:

- experiences significant delays at the IPv6 layer of the network stack on handoffs away from the home network.
- remains agnostic of the link-layer mechanism at the cost of a significant L2 handoff delay component and makes no provisions that can alleviate such delay component, characterising the total MIPv6 handoff delay measure.

3.9.2 Statistical Distributions for Delay Components

In this chapter we have also derived a number of statistical distributions for the individual delay components experienced during a MIPv6 handoff. This set of statistical distributions are used in subsequent simulations to describe stochastically the statistical measure of MIPv6 handoff delay during simulations, when compared with any novel IPv6 mobility management proposed in later sections.

3.9.3 Experimental evidence vs. MIPv6 specification claims

From the above experimental evidence, we can deduce that true MIPv6 handoff performance is conflicting with performance claims made by Mobile IPv6 specification [32]. In particular, Mobile IPv6 *cannot* guarantee a handoff rate of 1 handoff/sec. This arises from our demonstration that a MIPv6 handoff away to a visited network (h2v) is guaranteed to last a minimum of 2.8 sec. Our results show that the Mobile IPv6 specification must readjust its claims to a handoff every 3 seconds, without including any influence for network externalities such as RTT disparity over Internet links.

Furthermore, a high rate of router advertisements (small RtAdv interval) by itself *does not* assure a faster handoff. We have shown that in best cases a minimum average of 80ms of hangover delay arises from the moment the L2-handoff has completed until any subsequent neighbour discover function. The fundamental reason is that movement detection on the MN relies two possible time-based events: (i) the expiry of the router advertisement lifetime (ii) untimely initiation of router solicitation. With respect to

[22] if the context is required before the MN is allowed access to the visited network

the measure of hangover delay we have acknoledged that two factors are potentially responsible for its introduction: (a) lack of a standard form of movement detection that can identify fast MN's detachment from the current PoA as well as attachment to new PoA; (b) inefficiencies with Router Advertisment timers that are responsible for the implementation of the movement detection function at the MN.

3.9.4 On the size of WLAN domain deployment

Looking at the MIPv6 handoff performance and reflecting upon the enormous popularity and subsequent deployment of WLAN networks [222, 223]

In addition, the size of WLAN domains from the perspective of IPv6 handoff management, can be argued that it can only be small given the fact that:

1. WLAN cells are expected to maintain a small Tx range/AP for urban deployment scenarios due to frequent obstructions in their line of sight (buildings.

2. Due to the fact that WLAN cells operate in the unlicensed ISM band, there is reason to expect high deployment competition amongst many potential WLAN WISPs within a densely populated geographical area.

3. Due to the small Tx range of WLAN cells it is essential and expected that a geographical area requires high 802.11b/g AP density for complete coverage. However, WISP cannot be expected to install such APs at locations of their preference; this would incur a very high cost in location acquisition or leasing for AP installation. As a result each WISP would only be able to afford small domains, considering the high demand by competing WISPs for AP installable locations and their deployment costs.

As a result of small WLAN ISP domains and given small Tx range/cell, MNs would be expected to experience a frequent inter-domain IP handoff. This is further augmented by parallel deployment of WLAN ISP domains in the *same* geographical location, i.e. competing WLAN overlays. The former implies that, the rate of IPv handoff is expected to increase simply by the prospective reality of future WLAN deployment scenarios, manifested as frequent inter-domain IP handoffs, with less sensitivity on the mobility pattern (speed or direction) of the MN. Under such scenarios MIPv6 handoffs, are bound to paralyse any notion of interactivity during VoIP communications between the MN and its peers.

A number of interesting conclusions arise from the above experimental results of MIPv6 handoff performance which if combined with worst case RTT delay, have got a cascading effect upon the commercial future of IPv6 mobility management currently realised through the MIPv6 standard.

3.10 Conclusions

This chapter presented an in-depth experimental analysis of the MIPv6 handoff process. In particular, we have shown that a MIPv6 handoff effected from the home/visited to a visited network (h2v), experienced a delay of nearly 2.7sec; out of this delay figure around 2.38 seconds are attributed to MIPv6 handoff delay incurred at the network layer. The former validates that *the MIPv6 handoff process cannot support delay-seamlessness for the purposes of interactive real-time services.* For the case of a MIPv6 handoff conducted from the visited to the home network (v2h), the total MIPv6 handoff delay reduces down to 406ms of which on average only 20ms are attributable to the MIPv6 handoff process at the network layer.

For the case of the v2h MIPv6 handoff, *the diminished measure of delay performance is not due to the performance of the MIPv6 protocol itself but to the co-location of the HA and the MIPv6-agnostic AR onto a single network device.* In such case the MIPv6 function is assisted by state maintained by the HA, while dealing with reachability and link access through the AR component of the device (which in practise is HA/MIPV6 unaware).

For the case of an h2v MIPv6 handoff, *the excessive measure of MIPv6 handoff delay comprises of multiple latency components*; these are incurred by four fundamental factors presented with order of significance: (i) *Duplicate Address Detection delay,* (ii) *Neighbour Unreachability detection delay,* (iii) *Hangover delay,* as a result of movement detection inefficiencies (iv) *L2-handoff delay.*

These handoff functions exhibit, in the h2v type of MIPv6 handoff, the respective average delay contributions: (i) 1001ms (ii) 1010ms (iii) 53-404ms depending on the type of hangover delay (iv) 423ms for the particular 802.11b vendor implementation.

The case of a h2v MIPv6 handoff confirms further that *the MIPv6 handoff standard cannot support a handoff rate of 1 handoff/sec, since clearly within that period a handoff has not completed.* The true handoff rate that the specification can support, according to our experimental results appears to be 0.35 handoffs/sec for the *h2v* case and 2.46 handoffs/sec for the *v2h* case.

All of above derivations are justified by *the reactivity of control signalling in the face of MN's detachment from its current PoA.* This is clearly evident by observing the delay performance of the v2h type of MIPv6 handoff; as a result of active state maintained by the co-located HA/AR, the MN does not consume a significant period in either DAD or NUD function. Furthermore, as a result of such state existent on the co-located HA/AR system hangover delays are eliminated, indicating efficiency in neighbour cache management by exploiting the active entries of the defending HA mobility management entity.

The above are confirmed by the first statistical moments for the v2h cast of MIPv6 handoff; there the average delay total comprises of the following factors presented in order of significance: (i) L2-handoff delay (ii) NUD delay (iii) DAD delay. These handoff functions exhibit the following average delay contributions: (i) 376ms (ii) 16ms (iii) 13ms.

The above measures of delay have a cascading effect on the measure of delay variance experienced by the communicated flow. In particular, for the case of the h2v MIPv6 handoff jitter has an average value of 175ms, with the v2h case reporting an average jitter of $< 30ms$. Similarly the jitter amortisation period for the h2v handoff case has a mean value of 1.7sec, whereas the v2h case has a mean value of 876ms. This implies that a MIPv6 handoff is guaranteed to destabilise lip synchronisation if effected with a rate faster than the aforementioned periods. Such possibility is of course pre-empted by the average measure of h2v handoff delay which is significantly greater ($2.8 >> 1.7$ *sec*).

We may note that *the above measure of delay accounts only for the simple case of IPv6 addressing and routing state establishment. In the event that the MIPv6 handoff process requires further AAA or QoS state establishment, for the MN at the new PoA, the delay measure becomes significantly more pronounced.* This is because AAA or QoS state if effected reactively is bound to require both addressing and routing at the MN *before* it can be established.

It is, thus, concluded that further optimisations to IPv6 handoff delay performance are necessary to support interactive and real-time IPv6 applications in a mobile context. The dominant delay contributions for the h2v type of the MIPv6 handoff process, arise from network-layer MIPv6 management signalling; as such they remain generalisable for any technology employing native IPv6/MIPv6 signalling.

It should be noted that the generality of results obtained from the behaviour of 802.11 link-layer handoffs, may be of limited use for non 802.11x wireless MAC/PHY technologies. This is because each wireless technology at hand provides particular benefits under specific design trade-offs. Such trade-offs deal with ease of use, deregulated frequencies, availability of bandwidth, cost of ownership as well as sheer openness in services over packet-switched wireless networks. To this end, we conclude that findings pertaining to L2-handoff performance remain technology-specific over the IEEE802.11 protocol specification.

Chapter 4

Seamless Multi-context IPv6 Handoff management

4.1 Introduction

Chapter 3 has demonstrated experimentally that Mobile IPv6 incurs significant delay during an IP handoff, independent of the router advertisement transmission rate. We have shown that the magnitude of latency of a single (h2v/v2v) MIPV6 handoff is large enough to impede any notion of interactivity or hard delay bounds of real-time delivery for VoIP flows between the MN and its peers.

It is intuitive that under MIPv6 or MIPv6-dependent handoff management mechanisms, an increasing IP handoff rate exacerbates such latency and, thus, any disruption of packet flow between the MN and its peers.

The above attest categorically that, delay seamlessness is a function that fails to be addressed by the recently ratified IPv6 mobility management standard. Seamlessness has been identified as the principle of strict adherence to delay bounds during MN's IP roaming between last-hop wireless infrastructure IPv6 networks; such delay bounds are essential for the purposes of interactive, real-time, packet communications of MN with its peers.

Preserving the *seamlessness principle* in best-effort wireless access IPv6 networks *WLAN or other*, emerges as an advanced IPv6 mobility management function with respect to existing IPv6 mobility standards. In this light, it is essential that next generation IP mobility management can allow robust handling of seamlessness by design, with reference to IPv6 *handoff delay.*

4.1.1 Multi-context IP state establishment

Seamlessness, however, may be applicable to multiple contexts of IPv6 mobility management, Operational deployment of MM mechanisms are expected to require authentication and billing (AAA) on the part of the service provider and minimum service quality guarantees (QoS) on the part of the user subscriber.

Hence, the generalised view of the seamlessness principle dictates that the above MM mechanisms must be *generic* enough to act as an efficient architectural substrate that foster *evolutionary* forms of IPv6 mobility management; that is, support sound mobility management foundations that are generic enough to support an *extensible* form of IP mobility management.

To this end, it is essential that efficient open forms of IP handoff management *preserve* delay-seamless performance, while supporting mobility control over *multiple contexts* of IP connectivity during an IPv6 handoff.

4.1.2 IP handoff selectivity

At the same time, the handoff management function should embrace information control, affecting *choice* in the IPv6 handoff decision pertaining to vertical or horizontal multi-AR-candidate IP handoffs. The latter arises from emerging perspectives of future commercial deployment scenarios [224, 225, 226]; these, envisage *availability* of multiple competing Wireless Internet Service Providers (WISPs) with a geographical area, offering Internet access over different technologies (GPRS or WLAN) with differentiated service, performance, or tariff *capabilities.*

The need for being selective during the IP handoff decision becomes clearer when considering the perspective of *nomadic* [70] IPv6 mobility. An MN-nomad is defined as the wireless host that is not attached permanently to, or limited, by any administrative network domain; its user is free to effect mobile IPv6 communications (or migrate indefinitely) over any wireless network domain meeting its travel path, willing to accommodating its traffic.

Currently, the MN remains agnostic to characteristics of different context, pertaining to IP connectivity, performance or even service *capabilities*, supported by visited networks. Such networks as part of a larger provisioning domain appear 'ad-hoc' in MN's transit path, as it roves between new points of attachment (PoAs) towards its destination. As a result, awareness of IP service capabilities either cannot be facilitated, or requires significant human intervention in discovery and configuration at the MN.

In addition, much like cellular networks, multiple wireless ISPs may be concurrently available. Hence, multi-domain IP service capabilities will soon populate the wireless ether in terms of service provisioning. The future mobile user is expected to be able to *choose* the 'best' of such capabilities, if he is to achieve an optimum of the best-effort IP-mobile network services available. Such optimum is defined according to his criteria or policies [227], whether personal or corporate.

Thus, in a competitive multi-WISP environment, efficient handoff management should also allow dynamic availability of context-aware information to the MN, pertaining to IP service or performance characteristics of its current or future PoAs, in support of IP handoff selectivity.

To this end, this chapter investigates efficient forms of handoff management by attacking concurrently the aforementioned three major challenges in support of advanced IPv6 mobility management: (i) delay seamlessness, (ii) multi-context state establishment (iii) handoff selectivity in aid of maximising MN's performance utility.

4.2 Problem Description

With respect to delay seamlessness, the measure of latency in IPv6 handoffs has been described in the previous chapter by two distinct components, identified by order of significance:

1. delay at the network layer: This is the total delay incurred by signalling at the network layer for the purposes of effecting a handoff between IP addresses.

2. delay at the link-layer: This is the total delay incurred by signalling at the link layer. For the purposes of this investigation the link-layer of choice has been the IEEE802.11b/g protocol (MAC and PHY sub-layers) family.

From the perspective of link-layer, Chapter 3 has shown that the latency induced by the link-layer handoff process can affect significantly real-time delivery guarantees of packet delivery to/from the MN. While such fact is acknowledged, it can be seen that subsequent analysis and evaluation of L2-handoff optimisations becomes, by necessity, technology-specific. The former indicates that fundamental assumptions about *layer independence* between the link and the network layer may be preserved only *if* each layer attempts to reduce handoff latency *independently.* To this end, this chapter focuses on reducing or eliminating handoff delay at the IP(v6) layer.

From the perspective of the network layer, the performance of a MIPv6 handoff effected in reaction to the discovery of a new network link confirms that *reactive acqui-*

sition of IP connectivity state at the new point of attachment, is insufficient to address transmission delay seamlessness during an IP handoff.

Packets sent towards the MN are lost *until* its IPv6 handoff has completed. It, thus, appears that *the amount of disruption in packet communications between the MN and its peers is dependent (and proportional) to the delay incurred by the MIPv6 handoff process*; this is insufficient to address transmission delay seamlessness during an IPv6 handoff.

In this chapter, we reconsider the entire approach of standard Mobile IPv6 established through reactive handoff management mechanisms, by encompassing delay seamlessness as its fundamental *architectural* requirement; we investigate a novel IPv6 mobility management protocol architecture that allows the IP mobility management task to sustain transmission delay seamlessness during an IP handoff.

The above is achieved by addressing both of the above deficiencies of MIPv6 mobility management *by-design*: (i) reduction or elimination of delay incurred as a result of any factors pertaining to the initiation and completion of a handoff at the (IPv6) network layer; (ii) sustaining packet transmissions towards the MN by decoupling its dependency on the completion of the IPv6 handoff process. These deficiencies translate effectively to the requirement of delay-efficient forms of *handoff* and *flow forwarding* management as integral part of an advanced IPv6 mobility management architecture. In this light, this chapter focuses onto seamless forms of handoff management. The component of delay-efficient IPv6 flow forwarding management is presented in Chapter 5.

4.2.1 Hypothesis

The proposed IPv6 mobility management architecture advocates the need for *proactivity* in IPv6 handoff management signalling; such signalling entails *in-advance* manipulation of IP connectivity state at the new point of IP attachment. W argue *that a proactive IPv6 mobility management mechanism is capable of addressing transmission delay seamlessness* by reducing significantly (or eliminating) handoff delay at the network layer, while sustaining packet transmissions towards a reachable MN.

The above is investigated by first looking at the process IPv6 address configuration. IPv6 address configuration is identified as a form of IP state establishment, associated with the most critical context of IP connectivity, for packet communications between the MN and its peers: *identification* and *routing*; this is collectively defined as *IP Roaming State* context.

By identifying this core type of context state, the proposed mobility management architecture sets the foundations towards *generalising* the process of state establishment to *other* contexts pertaining to IP connectivity of the MN, such as Accounting/Authorisation/Authentication (AAA) and Quality of Service (QoS). This generalisation argues in favour of delay seamlessness over multi-context state establishment, essential for the initiation and completion of MN's network-layer handoff.

While the merits of such IP mobility architecture over different contexts of IPv6 connectivity state are elaborated, this study does not attempt explicitly to demonstrate its performance over multiple forms of state establishment, relevant to the delay performance of an IPv6 handoff. Instead, we present and analyse the measure of handoff delay performance attained through the principle of proactivity over a single but critical type of state context, that of IP Roaming state establishment. Similar handoff performance gains can be attained by applying the same architectural approach in parallel, over any other state contexts of IP connectivity.

With respect to handoff management, the proposed IPv6 mobility architecture presents further, the performance benefits incurred by enabling IPv6 handoff selectivity at the MN. By means of proactive signalling, the MN can select between candidate IPv6 handoff points of IP attachment, that increase its measure of service utility, subject to MN's satisfiable performance criteria or policies. In this manner, this study demonstrates that *by selecting the PoA that best fits its performance requirements during an IP handoff, the MN can increase its perceived measure of service utility/benefit, while sustaining IP communications with its peers on the move.* The type as well as level of such utility is specific to the utility function profile or criteria of the individual MN.

4.2.2 Outline

Section 4.3 presents related work pertaining to IP mobility management mechanisms aiming to improve IP handoff performance as well as associated trade-off weaknesses.

Section 4.4 identifies a set of requirements for the design of robust IPv6 mobility management supporting seamlessness proactively.

Section 4.5 presents the architectural rational of proactivity semantics in IPv6 mobility management. Section 4.6 presents the architectural core of the proposed Proactive IPv6 mobility management architecture with subsequent focus on IPv handoff management.

Section 4.7 presents the core algorithms as well as a protocol mechanism supporting dynamic discovery of handoff PoA neighbours, within the mobility path of the MN.

This achieved by exploiting the natural adjacency of coverage areas to form a mobility neighbourhood grouping. Such grouping is then mapped onto the respective set of ARs, each mapping to its own point of attachment (PoA).

Section 4.8 presents the manner in which state pertaining to the IP connectivity of the MN, or capabilities of the candidate ARs, are established or relocated in-advance of its next IPv6 handoff. It demonstrates its application by means of establishing proactively critical state (IP-Roaming), for the achievement of an IPv6 handoff. Subsequent parts of this section describe the coupling of state establishment with proactive handoff management, including the management of PoA neighbour membership before, during and after the completion of a proactive IPv6 handoff.

Section 4.9 evaluates the performance of the proposed proactive handoff management model by means of discrete event simulations; it identifies the measure of IPv6 handoff delay, jitter and packet loss over both reactive and proactive MIPv6 handoff management mechanisms. These metrics are derived over a varying measure of MN speed and pause period, tracking important aspects of non-determinism in the mobility pattern of the MN.

Furthermore, our simulation study investigates the influence of state convergence in handoff AR discovery onto proactive IPv6 handoff performance over the above parameters, with respect to varying MN and PoA densities.

The last part of this performance assessment investigates the maximisation of MN's utility during IPv6 handoffs, as a result of IP handoff selectivity. The measure of service handoff utility is evaluated by employing proactive availability of PoA handoff diversity information, during an IP handoff and contrasted against traditional SNR handoffs effected normally over Mobile IPv6. Section 4.10 presents results derived from the simulation study on the above assessment objectives.

Section 4.12 presents a summary of our findings with conclusions on proactive IPv6 handoff delay performance without support of flow forwarding management.

4.3 Related Work

Snoeren et al, proposes an end-to-end IP mobility mechanism based on dynamic naming (DNS updates) [168]. Under this mechanism, an IP network transition requires the MN to obtain a new IP address and update the DNS mapping for its host name. While the implementation approach demonstrates the feasibility of the such a mechanism, it remains unsuitable for delay-bounded multimedia applications, due to DNS-update latencies.

Several other approaches have been proposed for managing micro-mobility independent of the IP protocol family. Micro-mobility management employs typically a single gateway per network domain empowered with the task of maintaining a routing database that maps host identifiers into their current routing location. Amongst them Cellular IP [228] and HAWAII [229]; these approaches are identified (see Chapter 2) as routing-based, localised micro-mobility schemes, in support of improved signalling and under certain cases handoff delay performance. Both follow a common architectural strategy, whereby a domain gateway registers its address with the HA and forwards packets to the MN's home address used within the domain.

Cellular IP requires a legacy routing mechanism replacing IP routing within a network infrastructure. On the contrary, HAWAII does not replace IP routing but requires significant extensions to it, since dynamic routing is replace by mobile-specific (aka host) routes within a network domain. Despite their appeal, these mechanisms encounter significant trade-off limitations of network scalability and extensive modification to the existing IP infrastructure versus localisation of latency and location update signalling. However, they provide significant insight in terms of performance alternatives in localised mobility management. They show that despite scalability limitations, path rerouting can be effected efficiently by identifying crossover points between the previous and new route of the MN as long as fault tolerant routing and its cost can be sustained.

Similar hierarchical mobility approaches have been followed by the Mobile IPv6 *Regional Registrations* approach [230] whereby a routing hierarchy is created on local mobility agents co-located with network routers within a local domain. In a fashion similar to HAWAII, this mechanism is significantly dependent on multi-level routing hierarchies acting as an additional routing overlay in all network domain routers, in addition to dynamic routes maintained by IP. In addition, it prescribes the use of multiple tunnels between routing hops that introduce processing complexity in the intra-domain routing function

Hierarchical MIPv6 [30] removes the dependency of routing hierarchies within a domain by employing a single forwarding point between the edge of the network and MN's point of attachment; this is effected by means of a tunnel emanating from mobility anchor point (MAP) router terminating at the MN.

In the realm of hierarchical mobility management mechanisms belongs also the

IDMP [231] and Multicast Mobility [232] proposals. The main difference between these mechanisms and Hierarchical MIPv6 is the use of multicast for the purposes of intra-domain flow forwarding at the edge of the network domain.

With the exception of Cellular IP, all aforementioned micro-mobility management proposals do not address latency as a result of *state configuration and establishment*, for state such as IPv6 addressing and routing. Cellular IP can afford such capability at the cost of changing the *entire* routing function, such that the MN requires *only* its home IPv6 address. Such effects apply also in the case of multi-context state establishment for IP handoff purposes.

In addition, all hierarchical mobility management mechanisms emphasise their limitation in terms of domain gateway failures. It can be seen that in the event of a failure at a domain mobility gateway, any of the above localised mobility management mechanisms is guaranteed to collapse; in such case, IPv6 mobility management services fail for *all* MNs within a domain.

Furthermore, significant complexity is introduced in the event of multi-gateway domain support at the edges of the network, both in terms of *configuration* and *routing*; configuration issues stem from the challenge of apportioning the correct domain segments to a single domain gateway with failure resiliency capabilities. Routing issues stem from the non-determinism of downstream arrival of a flow destined to the MN. In the case of large (multi-border) network domains, while the MN may register with a domain mobility gateway G_i, there is no guarantee that the downstream flow directed towards the MN will enter the domain from that particular gateway. As a result either suboptimal routing would emerge where the flow has to be redirected, or the flow cannot be routed towards the MN. To date there exists no such robust configuration mechanism that can sustain the survivability of a localised mobility management mechanism and afford to overcome efficiently its scalability limitations.

Fast Mobile IPv6 has been a recent research effort in the IETF aiming to minimise the handoff latency of MIPv6 [33]. Its handoff management function requires link-layer information to aid the configuration of the care-of address of the MN on the new AR before MN upcoming IPv6 handoff. The emerging limitation in this proposal is that the IP handoff management function is dependent critically on link-layer detection functions at the MN, to determine PoA handoff neighbours, while in active IP communications with its peers. The accuracy of such information is highly dependent on increased signalling (scanning) overheads at the link layer, that increases non-linearly within a single coverage area. Chapter 6 investigates further the performance of FMIPv6 against the performance of the proposed mobility management architecture.

Furthermore, as seen in Chapter 2, emerging results from independent investigations on FMIPv6 performance, report prohibitively large handoff delays as a result of an increased number of wireless hosts attached to a wireless link. This is because latency from FMIPv6 network layer *signalling* becomes dependent on access contention before signals can be propagated between previous and new ARs.

4.4 Requirements for next-generation IPv6 Mobility management

Investigations conducted, thus far, in [46, 44] and Chapter 2, have revealed a number of performance as well as functional limitations emerging from first-generation IPv6 mobility protocol proposals, such as Mobile IPv6 and/or IPv6 localised (micro/hierarchical) mobility management mechanisms.

Furthermore, experimental measurements of Chapter 3 have identified factors that can affect significantly the delay incurred by the IPv6 handoff process and as a result seamless handoff performance. In addition, reported results on the performance of emerging second-generation mobility management proposals, such as FMIPv6, expose further performance limitations with respect to handoff delay.

It becomes apparent that a robust IPv6 mobility management architecture that addresses successfully limitations of previous mobility management mechanisms, including delay seamlessness, must encompass the set of these findings as part of its underlying system design. To this end, this study consolidates identified experimental results and findings from both personal and independent investigations, into a requirements specification for *any* next-generation IPv6 mobility management. It is important that such requirements are considered both collectively as well as individually, since many of these requirements are interrelated.

Such specification encompasses the following set of IPv6 mobility management requirements in support of advanced IP mobility management (MM):

- IPv6 MM must be able to support *delay transparent* handoffs at the network layer. Such support cannot account for delay incurred either as an external factor or for factors that are beyond the control of the network layer.

- IPv6 MM must not rely solely on the state available at the network layer, if it is to support efficiently delay-transparency IP handoffs. It is essential that it exploits a controlled set of generic link-layer state information in the form of link-layer (L2) *triggers*. This is justified by the fact that an L2-handoff precedes the actual IP handoff; as such the link-layer must notify immediately the network layer of such events.

- IPv6 MM must be able to sustain delay seamlessness in an IPv6 handoff in the face of *multi-context state establishment* pertaining to the IP connectivity requirements of the MN at the new point of attachment.

- IPv6 MM should be able to allow *handoff selectivity* in the event of multiple new handoff candidate points of IPv6 attachment. Currently, no IPv6 mobility management can provide such capability to the MN at the network layer

- IPv6 MM must sustain *layer independence*; that is, remain *independent* of the underlying link-layer technology.

- IPv6 MM must not introduce dependencies in the core routing function to support the IPv6 mobility of the MN. Such requirement refers to the routing core of a single network domain as opposed to access routers as the point of IPv6 attachment of the MN.

- IPv6 MM must sustain *distributed reliability* by not introducing critical single points of failure. Localised mobility management schemes have demonstrated the limitations from the tendency for natural concentration of control over a single mobility gateway. They have also demonstrated the increased complexity induced by extended requirements for mobility gateway failure and configuration.

- IPv6 MM must remain *scalable*. Scalability is defined as the ability of the IPv6 MM signalling to sustain near-linear growth for realistic measures of increase in MN population size.

- IPv6 MM should sustain route optimality during its handoff or flow forwarding deliberations. *Route optimality* is defined as the ratio of the routing cost from an IPv6 MM mechanism over the routing cost incurred by a straight unicast path between the MN and its peers while at the new point of attachment.

- IPv6 MM signalling must account for wireless link contention; on this basis it must strive to minimise delay influence from access contention instigated by the wireless link layer.

- IPv6 MM signalling should minimise signalling overheads relative to the benefit of the function effected. Clearly, additional functionality, (if essential) is expected to incur more signalling overheads than the lack of its support.
- IPv6 MM must support both forms of horizontal as well as vertical handoffs, in a delay-efficient manner.
- IPv6 MM must support non-repudiated signalling in its control deliberations.
- IPv6 MM must not depend on standard link-local IPv6 signalling, such as Router Advertisements, for the purposes of Movement Detection. Such reliance:
    1. introduces an increase of signalling overheads over the wireless link. Such overhead consumes essential bandwidth over the wireless link, while it amplifies MAC contention from a single transmitting source.
    2. is dependent on isochronous transmission interval of the control signal. For increasing/dense node population under a single AP, such isochrony cannot be guaranteed, resulting a delayed identification of the new network at the MN. This has the cascading effect of delayed IPv6 handoff completion and subsequently latencies that affect adversely the performance of interactive communications.

With these requirements in mind this study embarks towards the design of a novel proactive IPv6 mobility management architecture, supporting delay seamlessness in the handoff management function of MN's IP mobility.

4.5 Towards seamless IPv6 Mobility Management

Chapter 3 has shown that from a handoff management perspective, certain MIPv6 functions appear to be delay-intensive for interactive real-time purposes. More accurately, *the form* in which MIPv6 employs particular core IPv6 protocol functions, for the purposes of movement detection, duplicate address detection or neighbour reachability has been shown to incur significant delay in the overall IPv6 handoff process.

It is important to note in the course of introducing novel protocol semantics, such as IPv6 mobility, every effort must be made to preserve, where possible, existing core IPv6 protocol functions. Significant changes to existing operational IPv6 protocols required by new protocol mechanisms, impose considerable implementation changes to the existing IP protocol stack for both existing host and router devices. This has got a ripple effect on established forms of IP communications leading to stagnant incremental evolution in Internet protocols.

As a result, novel protocol semantics meet higher acceptance from the Internet community, when they have a minimal (change) impact on the existing IP protocol base, while remaining efficient. As seen in Chapter 2, such is the case with the semantics of Mobile IPv6 and its operation, complemented by existing core protocol functions of the IPv6 protocol suite.

While MIPv6 affects minimally existing core IPv6 protocols, it employs them *reactively* during MN's IPv6 handoff. A mobile host performs all necessary functions pertaining to an IPv6 handoff, such as CoA address configuration or routing, *after* its has detached from its previous point of attachment. Essentially, these IPv6 handoff functions prescribe the establishment of IP Roaming state through the use of Neighbour Discovery [107]. Hence, IPv6 Roaming state is established *after* an IPv6 handoff has been initiated.

Reactive utilisation of IPv6 Neighbour Discovery functions, by the MIPv6 handoff process, becomes one of the fundamental reasons for increased handoff delay, that impedes any adherence to guarantees for real-time delivery of IP traffic.

To support real-time delay seamlessness while preserving existing IPv6 Neighbour Discovery functions, this study asserts that IPv6 handoff management must be effected *proactively.*

4.5.1 Semantics of Proactivity

In an abstract framework, this study identifies proactivity as:

Definition 4.1 *the form of forward stimuli*[1] *cognition effected from previous learning, that remains active in a subsequent activity.*

That is, proactivity is a form of information or event manipulation that acts in advance of future activity. The above definition may be applicable in *any* form of *event* management: within a proactive environment, an entity -whether protocol, policy or other- may *create* or *influence* one or more future events, by acting in *preparation* or *anticipation* of these events, through learning.

It may be seen that the above definition differentiates between the notion of preparation and anticipation. Preparation encompasses no indication of a tentative time

[1] signalled

deadline about the occurrence of the future event. Anticipation encompasses some form of indication of a tentative time deadline on event occurrence by means of predictive techniques. A proactive event, may either be regenerative in terms of event instances or simply influence the course of events as they occur.

Notwithstanding, in all above cases, proactivity is effected so as to *enhance* future activity. In this light, it may be viewed as an intelligent form of forward event or data manipulation.

4.5.2 Proactive Mobile IPv6

Applying the above semantics in the context of IP mobility, allows to establish a model that promotes proactivity in IPv6 mobility management mechanisms. Such IP mobility model emphasises that advanced IPv6 handoff management cannot rely on reactive state manipulation at the new point of IP attachment; this is far too slow, to real-time interactivity purposes; instead, the network must *proactively* manage and distribute a mobile node's IP connectivity state much in advance of the MN's upcoming IPv6 handoff transition. In this manner, mobility management at the network layer, can support real-time delivery/transmission guarantees for active IP flows to/from a mobile host.

Proactivity, however, is based on forward learning; that is, advance gathering of information pertinent to some future activity. For IPv6 mobility purposes, this means *advance* gathering of all information essential for MN's future IPv6 handoff. The *composition* of such information is structured as state material to IP connectivity. The *type* of such information is clearly context-dependent: different information must be collected for different types of IP connectivity state, such as IP Roaming, AAA or QoS.

Nonetheless, besides the composition or type of information that must be collected in advance of MN's IPv6 handoff, the *source(s)* of such information must first be identified. Since IP mobility is concerned with host movement at the network layer, the primary source of IP connectivity state, independent of the context, is identified to be the *access router (AR)*. It becomes, thus, imperative that the access router involved in MN's next network transition, is *identified* and *enquired* about IP connectivity state, essential to MN's next IPv6 handoff.

It can be seen that for an increasing network domain size, identification and management of IP connectivity state as well as their sources cannot be afforded manually. For administration purposes, it becomes infeasible to afford manual discovery and configuration of access routers supporting IPv6 handoffs, particularly within growing network domains that foster *incremental deployment.* As such it becomes unrealistic to expect manual exchange and maintenance of IP connectivity state information, supported by access routers that participate accommodate MN's IPv6 handoff.

Since the MN is agnostic of the network that is transiting to during an IP handoff, forward learning of IP handoff-relevant state information will naturally require *cooperation* between network domains. Hence:

Definition 4.2 *proactivity is based on some minimal form of cooperation; otherwise forward data learning becomes unattainable.*

The above implies that in the event of non-cooperating network domains, proactive IPv6 handoff management is infeasible.

Drawing from the analog of cellular networks, it can be seen that despite competition, cooperation between cellular network providers prevails. It is manifested to the mobile subscriber, through *provider roaming agreements*, as network *roaming* capabilities. In fact, cooperation becomes essential for any wireless network provider, since: (i) it provides redundancy in the event of network failure, (ii) provides geographical span in service coverage. This is why cellular communication becomes widely accepted as a ubiquitous service in forms similar to plain telephone services or roads. For IP networks, cooperation is one of the fundamental design requirements: packet routing is based on cooperative forwarding.

From the above it can be seen that for the purposes of IPv6 mobility management, proactivity is a viable architectural approach, that in fact strengthens the notion of IP network cooperation for other IP communication protocols such as IP Multicast. This is because, in wireless access IP networks, cooperation between network domains increases the shared economic utility among all underlying service providers; each can offer (and controls) a distinct part of the total geographical coverage. We postpone any further elaboration on this issue until Chapter 5.

Having established the basis for proactivity in IPv6 handoff management, it becomes essential to identify a mechanism that can: (i) identify access routers (ARs) that emerge as candidates in accommodating MN's next IPv6 handoff, (ii) identify and establish forms of forward state manipulation pertinent to MN's IP connectivity in view of its next IPv6 handoff.

4.5.3 Proactive versus Reactive IPv6 Handoff management

It is important to note that Proactive MIPv6 is distinctly different from the existing reactive MIPv6 standard [32]. Standard MIPv6 cannot be extended to effect an IPv6 handoff proactively, while eliminating the disruption of the IPv6 flow towards the MN, *without* changing its fundamental design; a MIPv6 handoff is performed by-design reactively; it extends existing Neighbour Discovery mechanisms to support movement detection, address and router configuration *after* the IPv6 handoff has been initiated. This becomes apparent when the proposed mobility management architecture is contrasted against the MIPv6 signalling requirements, analysed in the Chapter 3.

To the best of our knowledge, the majority of the proposed mobility management mechanisms such as MIPv6, CIP, HAWAII, HMIPv6, or FMIPv6, remain agnostic of the set of ARs that are candidate to accommodate MN's next IPv6 handoff. With the exception of FMIPv6 and CIP, for all above protocols the MN must first transit to the new network, receive a new (solicited) router advertisement and/or react to the former through address resolution/DAD/Neighbour reachability or lack of it. This makes obvious that reactive discovery of the new AR upon completion of the L2-handoff, acts to the detriment of fast completion of the IPv6 handoff.

FMIPv6 in particular, makes an implicit assumption, during a 'make-before-break' handoff scenario that the current AR has some *a priori* knowledge concerning the next AR; under FMIPv6, the current AR issues proxy router advertisements for a neighbouring AR; however, FMIPv6 has no explicit knowledge of this AR, nor does it make any protocol provisions for discovering it.

Furthermore, for the purposes of vertical handoffs over wireless technology hybrids, most IP handoff management proposals assume a very strict ordering on the layering of the networks. MNs are assumed to know *a priori* which of the wireless (technology) layers are available and have a 'hard-wired' precedence for each network based on (fixed) maximum bit-rate offered by the underlying wireless technology [227].

Recently, such need has also been acknowledged in the IETF. To this end, the Seamless Mobility (SEAMOBY) Working Group has given rise to emerging protocol abstractions such as Context Transfers (CT) [233] or Candidate Access Router Discovery (CARD) [234], currently under investigation[2]. SEAMOBY has recently identified issues alleviated by future CARD and CT mechanisms [235], giving rise to respective requirement specifications standing as the basis of future CARD or CT protocol proposals.

[2]during the course of this investigation

At the time of writing, both CARD and CT draft recommendations are currently the subject of on-going research that has not yet matured towards protocol standardisation.

4.6 Proactive IPv6 Handoff Management

It is of paramount importance for any advanced IPv6 mobility management protocol to attain explicit MN addressing and routing information at the next AR, *in advance*, of the host's IPv6 handoff. Given that the mobility pattern or exact transit path of the MN is *not* known in advance of its next IP subnet roaming, accurate discovery of the next AR at *all* times appears to be realistically difficult, particularly in the face of ping-pong effects [236]. For this reason the requirement of obtaining information for the exact next AR must be relaxed to the one of acquiring such information for a *set of ARs* that are candidates for the MN's next IPv6 handoff.

The Proactive IPv6 mobility management architecture proposed herein, departs distinctly from the aforementioned schemes by identifying first a discovery mechanism for ARs that are neighbours within MN's *next mobility-hop* manifested during MN's IPv6 handoff [50]. Once such mechanism is in place, additional functions, tightly coupled to the discovery of candidate ARs, proceed to exploit identification and routing information of neighbouring points of attachment; in this manner core IPv6 Neighbour Discovery is expedited for the purposes of a seamless IPv6 handoff.

4.6.1 Architectural Substrate

Proactive IPv6 mobility management comprises of two basic functions essential for the purposes of MN's IPv6 handoff: (i) network layer handoff management (ii) flow forwarding management. Proactive IPv6 handoff management is in charge of forward state manipulations *in-advance* of MN's next IPv6 handoff. Flow forwarding management is responsible for reducing/eliminating packet flow disruptions towards the MN destination, *during* its IPv6 handoff.

For these to functions to behave proactively, however, it is essential that state has first been identified and established with *entities of interest*, in advance of any subsequent manipulations. For mobility management purposes the entities of interest are the network-layer (L3) devices allowing access to a provisioning domain over some wireless access point. Such L3 devices are known as *Access Routers (AR)*. However, wireless access is supported through an *Access Point (AP)* (a.k.a Base Station); a wireless AP implements the bridge between a wireless technology enabling mobile access and the access network through an AR.

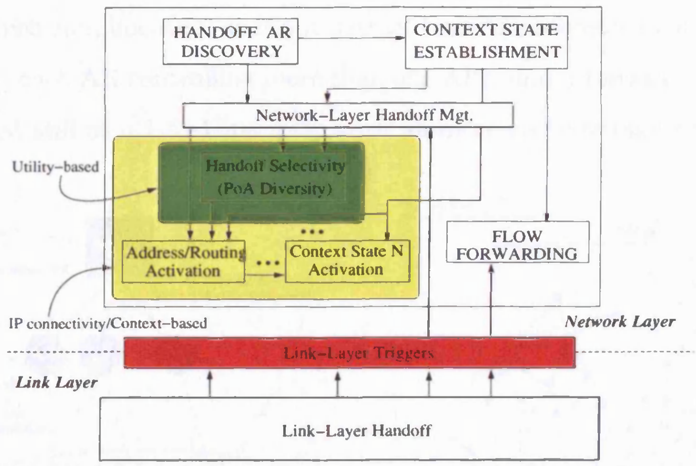


Figure 4.1: Proactive IPv6 Mobility Management Architecture

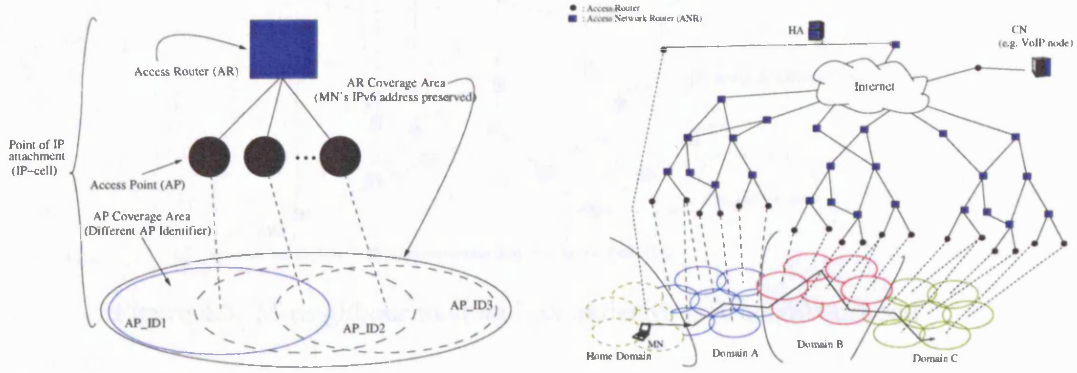
Thus, to enable identification and establishment of state pertinent to the IPv6 mobility management of an MN, it is essential that interested parties, namely a configuration of ARs and APs, are first *discovered*. Such requirement fosters the need for a *handoff AR-AP mapping* discovery mechanism; this is a function that feeds proactively a discovered AR/AP mapping into the mobility management function, to allow state establishment in mobility-management contexts of interest.

The above calls further for a context state establishment mechanism that explores the discovery information to enable forward state generation, pertinent to MN's next IPv6 handoff. Figure 4.1 presents a schematic overview of the key functions comprising the proactive IPV6 mobility management architecture.

Both flow forwarding and network-layer handoff management functions, employ an interface with the link-layer handoff of the MN. Reason for that is the fact that any network layer handoff is *preceded* by a link-layer handoff. Proactive mobility management [129] takes on the approach that by exploiting a minimal, generic set of link-layer information, it can support delay seamlessness in MN's next IPv6 handoff. The L2-triggers interface reports directly to the network-layer handoff and forwarding manager, allowing independence from standard IPv6 Neighbour Discovery signalling and delay associated with it.

Under the proposed mobility model an AR may be associated with one or more APs. While each AP is described by means of a different identifier, the AR together with the associated APs identify a unique *Point of Attachment (PoA)* as shown in figure 4.2(a).

One or more such PoAs define a wireless Internet Service provisioning (WISP) domain 4.2(b). Such simplification does not detract from the robustness of the mechanism since effectively each AR controlling more than one APs, under the same network prefix, can be modelled still as a 1-to-1 mapping with an increased coverage area.



(a) Point of Attachment (PoA) (b) PoAs within WISPs

Figure 4.2: Constituents of a Point of IP Attachment (PoA) as the last-hop extension of infrastructure WISPs

Bandwidth resources are expected to be constrained over wireless links; for this reason proactive IPv6 handoff management emphasises on *minimising signalling dependencies* over the air interface, particularly at the MN. This is because control signalling dependencies become performance-critical under harsh propagation conditions, as they typically result in increased bit error rates (BER) during MN's movement.

Each AR is expected to transit between three possible states: *new* (AR_n), *current* (AR_c) and *previous* (AR_p).

Each MN is assumed to associate or communicate at the link-layer with *one* AP at a time, before, during, or after an IPv6 handoff. This limitation is imposed to ensure that performance under Proactive IPv6 handoff management does not idealise link-layer access/association with multiple access points (or base stations). Idealised link-layer capabilities such as multi-AP access, afforded in techniques like multi-BS pilot signal detection [237] are not available across all wireless network technologies. For this purposes, this study makes a constrained assumption encompassing the lowest common denominator of wireless link-layer access.

Directly adjacent coverage areas are assumed to maintain a minimum measure of overlap with each other to allow continuous link-layer connectivity, and as such continuity in coverage. This prevents dis-connectivity as a result of lack of transmission coverage. Hence *the shape of the coverage area does not significantly influence the*

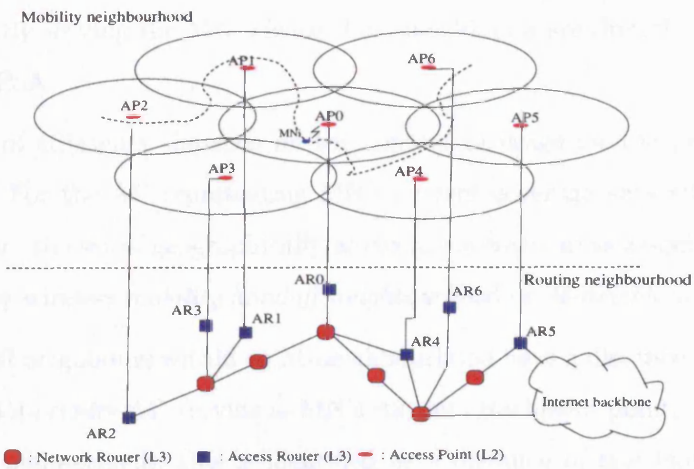


Figure 4.3: M-neighbourhood and its underlying R-neighbourhood

efficiency of proactive IPv6 mobility management from a BER perspective.

4.7 HARD: Handoff Access Routers Discovery

Proactive Mobile IPv6 initiates its handoff management deliberations by means of identifying the set of access routers that appear to be direct handoff candidates with respect to MN's current transmission range. This function is identified as *Handoff Access Router* discovery (HARD).

Under HARD, the first stage is to identify AR neighbourhoods supporting Proactive IPv6 handoff management before IP connectivity state exchanges can be established in preparation of MN's next IPv6 handoff.

4.7.1 Identifying Handoff AR Neighbourhoods

An MN roams, typically, between successive wireless coverage areas, manifested through APs, as it moves towards a destination. The geographical adjacency of these *coverage areas*, ensuring continuous access to the network, encompasses a logical separation between *geographical mobility* and IP network *routing*, in the current vicinity of the MN.

From the perspective of the MN, both of these contexts may be manipulated uniquely by means of a natural abstraction. This is the abstraction of handoff locality or *neighbourhood* around the *current* point of attachment of the MN. The idea of geographical locality in the vicinity of MN's handoff was first introduced in [238] for the purposes of cellular ATM handoff management.

Such abstraction is independent of the administrative domain. A handoff neighbourhood may be defined as the set of attachment points (PoAs), that are *adjacent* to the one currently serving the MN. Hence, PoA neighbours are directly reachable from MN's current PoA.

The type of adjacency depends on the context of usage for the particular neighbourhood set. For the AP representing MN's current coverage area within a wireless network domain, the set of geographically adjacent coverage areas associated with their APs identifies a wireless *mobility handoff neighbourhood* or *M-neighbourhood*.

AP handoff neighbours within an M-neighbourhood have a distance of one coverage area (CA) from its *centre* AP, serving as MN's current attachment point; under proactive IPv6 mobility management, this is identified as a distance of one *mobility-hop* from the current AP serving the MN and its immediately reachable AP neighbours; this is illustrated in Figure 4.3.

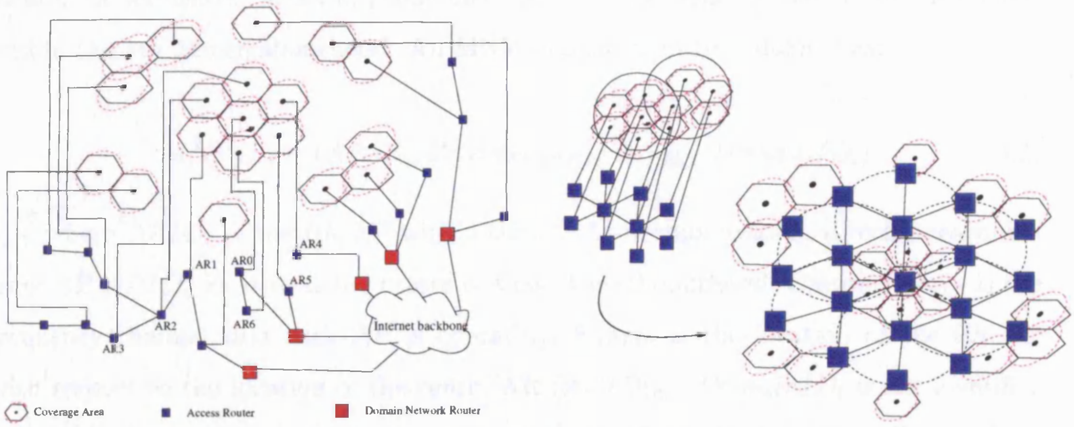
Thus, an M-neighbourhood is defined as the set of *geographically-adjacent* AP neighbours centred around MN's current point of attachment; the AP accommodating the MN is identified as the *current* AP while the surrounding APs are identified as AP *neighbours*.

Since each AP is expected to be attached to some AR, the set of all ARs associated with the CA constituents of the M-neighbourhood, is defined as a virtual *handoff routing neighbourhood* or *R-neighbourhood*, depicted in Figure 4.3. If the MN can perform a link-layer handoff between two AP neighbours of the M-neighbourhood, then the corresponding access routers are considered to be the primary (IPv6) handoff AR neighbours.

It may be seen that members of an R-neighbourhood are not necessarily adjacent within within the underlying network topology. Instead, the natural adjacency of coverage areas within the underlying M-neighbourhood mapping, provides an abstract, IP-mobility capable, virtual routing neighbourhood.

Hence, members of a handoff routing neighbourhood become mobility-hop adjacent, independent of the number of routing hops of their underlying network topology. Figures 4.4(a) and 4.4(b) show the transformation of the true network topology into a logical IP-mobility-aware topology mapping.

Thus, for the purposes of HARD identification, under proactive handoff management, a wireless network domain comprises of a set of IP-mobility-aware *M-neighbourhoods*; the availability of M-neighbourhoods is controlled by respective mappings of virtual *R-neighbourhoods* at the network layer.



(a) True network topology (b) mobility-aware topology transform

Figure 4.4: Transformation of true network topology into a logical network topology for the purposes of IP mobility management

Structured Handoff AR Neighbourhood membership

Members of a single M-neighbourhood are represented by its *M-neighbourhood vector (MNV)*. Such vector identifies the particular M-neighbourhood by means of the pair (MNV_k, θ_k) namely, its neighbourhood identifier MNV_k and its relative location θ_k with respect to adjacent M-neighbourhoods. The value of θ_k depends on the size of the M-neighbourhood emerging at the centre CA. Table 4.1 provides the respective angle of location for different M-neighbourhood sizes. The angle can be computed automatically upon determination of the size of M-neighbourhood.

M-Neighbourhood Size	degrees/neighbour
4	90
5	72
6	60
7	51
8	45
9	40
10	36
n	360/n

Table 4.1: variable-angle location dependent on the size of the M-neighbourhood

The MNV vector (MNV_k, θ_k) of the kth M-neighbourhood is defined as:

$$(MNV_k, \theta_k) = \{MNVe_{k,c}, MNVe_{k,1}, \ \dots \ , MNVe_{k,i}\} \tag{4.1}$$

where $MNVe_{k,c}$ identifies the *MNV-element* within the kth M-neighbourhood, located at its centre; $MNVe_{k,i}$ identifies the MNV-element of the ith AP neighbour within the kth M-neighbourhood. An MNV-element is in turn defined as:

$$MNVe_{k,i} = (APID_i, Channel_{APIDi}, \theta_{APIDi}, DomainID_l) \quad (4.2)$$

where $APID_i$ is the ith AP within the kth M-neighbourhood, directly reachable from AP $APID_c$ located at the centre of that M-neighbourhood; $Channel_{APIDi}$ is the frequency channel that such AP is operating; θ_{APIDi} is the location of the ith AP with respect to the location of the centre AR $(APIDi_c)$. $DomainID_l$ is the identifier describing the provisioning domain managing its set of attachment points PoA_i, where $PoA_i = (AP_i, AR_i)$. It important to note that the AP identifier and its operating channel act as the *best* discriminator, independent of the respective wireless technology.

Definition 4.3 *For each $APID_i$ within some (MNV_k, θ_k) there exists a unique $MNVe_{k,i}$.*

Definition 4.4 *For any two neighbouring access points $APID_i$ and $APID_{i+1}$, within the kth M-neighbourhood, the cardinality of their respective MNV vectors is not necessarily the same.*

The above two definitions allow us to describe each $APID_i$ with a unique MNV_{k,θ_k} identifier, while permitting any two neighbouring APs to have different M-neighbourhood size. This is because during an IPv6 handoff the MN may well encounter a new M-neighbourhood at AR_n with more members (and thus handoff candidates) than the M-neighbourhood of AR_c.

Definition 4.5 *The location angle θ of an M-neighbourhood MNV_k is identified by the location angle of its centre AP, $(APID_c)$.*

Definition 4.6 *A neighbouring AP, $(APID_i)$, that belongs to an M-neighbourhood MNV_k has a location angle relative to the centre AP, $(APID_c)$, of that MNV_k.*

The distinction between M-neighbourhood location and AP location provides a two-level granularity for MN location determination for the purposes of predictive handoff management under Proactive Mobile IPv6[3].

[3]Predictive handoff management is currently beyond the scope of this study and as such left as future work.

Similarly, the corresponding virtual R-neighbourhood mapping of MNV_k vector with respect to the current AR AR_c, serving the MN at the centre of the R-neighbourhood, defines the corresponding *R-neighbourhood vector (RNV)* denoted as:

$$\begin{aligned} RNV_k = \{&(MNVe_{k,c}, RNVe_{k,c}), \\ &(MNVe_{k,1}, RNVe_{k,1}), \\ &\vdots \\ &,(MNVe_{k,i}, RNVe_{k,i})\} \end{aligned} \tag{4.3}$$

$RNV_{k,c}$ identifies a unique *RNV-element* in a 1-to-1 mapping with the respective $MNV_{k,c}$ element at the centre of the M-neighbourhood. The $(MNV_{k,i}, RNV_{k,i})$ pair identifies the mapping between the ith MNV-element and it corresponding RNV-element within the kth joint M-R-neighbourhood respectively. It describes the ith attachment point PoA_i in the kth joint M-R-neighbourhood mapping. An RNV-element is defined as:

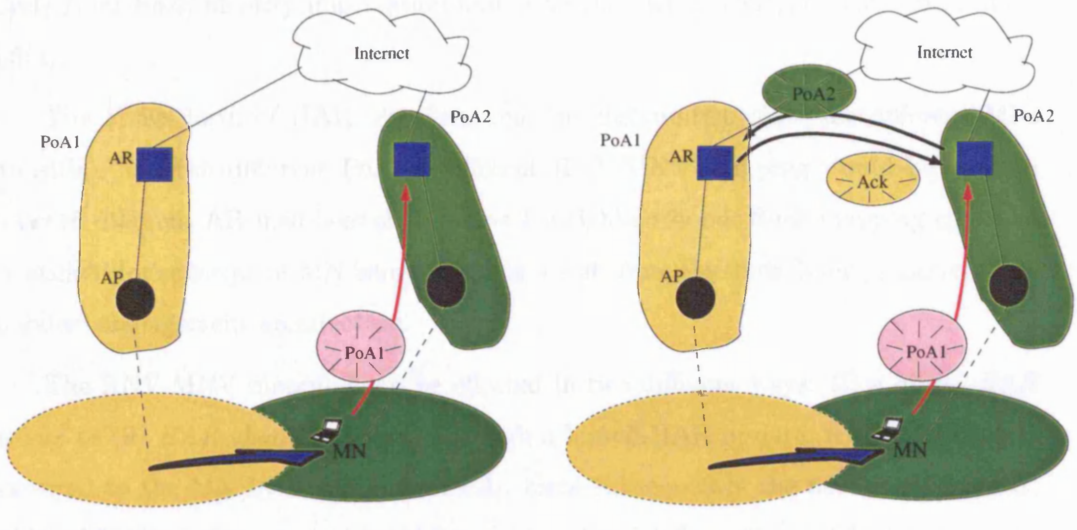
$$RNVe_{k,i} = ((IPaddr, PLen, LLA)_{k,i}) \tag{4.4}$$

where $IPaddr$ identifies the IPv6 address, $PLen$ the prefix length and LLA the link layer address of $AR_{k,i}$ associated with the ith AP within the kth M-neighbourhood. An R-neighbourhood maintains similar properties with the ones from definitions 4.3 and 4.4.

Each AR, as an integral component of the $PoA_{k,i}$ is configured with its own RNV-element and MNV-element, with each of them initialising its respective RNV and MNV vectors.

Both $MNV_{k,\theta_{k_i}}$ and its respective RNV_k, identify the *mobility-routing state* or *m-routing* exchanged between ARs in subsequent phases of handoff AR discovery; this is tracked by the availability of a MNV vector, effected over the corresponding R-neighbourhood. Exchange of m-routing state comprises of transmission of RNV vectors from every AR mapping to a unique CA member of the same MNV_k.

As a result, the handoff AR discovery process supports mobility-aware routing at the network layer while tracking MN's movement *omni-directionally.* In this manner, the availability of *m-routing* state can provide *next-mobility-hop* IP routing information within a M-neighbourhood, while the corresponding ARs are *not* necessarily next-hop



(a) Hinted-HAR update

(b) HAR-charged update: Hint+Handshake

Figure 4.5: Two forms of HAR identification under HARD: A HAR update solely by MN hints, (ii) a HAR update *charged* by an additional PoA handshake

reachable, in the underlying network topology.

Annex E.1 presents state optimisations relevant to a differentiated indoor/outdoor propagation environment, together with bootstrap learning dynamics for HAR discovery.

4.7.2 R-Neighbourhood Determination

As the MN moves towards some destination, it 'meets' new APs along its path, as shown in the M-neighbourhood part of Figure 4.3;

For access networks that support continuous coverage, it implies that the new AP associated is *adjacent*[4] with the previous AP encountered in MN's travel path and vice versa. For instance, AP_2 and AP_1 are found to be neighbours. Transition through adjacent APs associated with different ARs causes the MN to receive new router advertisements, indicating that a new point of attachment (PoA_{i+1}) has been encountered.

The 'discovery' of a new PoA by the MN, is possible to trigger a HAR neighbourhood update onto the new PoA_{i+1}, with identity information about MN's previous PoA_i. As a result, the underlying new AR (AR_n) can establish MN's previous AR (AR_p) as its own HAR neighbour. In this manner, an RNV-MNV vector mapping describing the two ARs for subsequent IPv6 handoff management, can be created effec-

[4]the term 'neighbouring' and 'adjacent' are assumed to have identical semantics and as such used interchangeably

tively from PoA_i identity information hinted by the MN to PoA_{i+1} as shown in figure 4.5(a).

The above form of HAR discovery remains distributed, since for different MNs transiting between different PoA, a different RNV-MNV mapping would be effected between different AR members of the same R-neighbourhood. Such mapping can then be utilised for subsequent MN handoff management, from the underlying proactive IPv6 mobility management architecture.

The RNV-MNV mapping can be effected in two different ways: (i) a *hinted-HAR update* or (ii) *HAR-charged update*. Through a hinted-HAR update, RNV information conveyed to the MN by its previous PoA_i, hints subsequently the new PoA_{i+1}; thus, a hinted-HAR update, provides HAR neighbourhood information solely by means of verified MN hints.

By means of a *HAR-charged update*, the hint provided by the MN is extended through an RNV-MNV handshake between the new PoA_{i+1} and previous PoA_i, as elucidated in figure 4.5(b). In effect, a HAR-charged update, expedites the HAR discovery process with a handshake from the new PoA_{i+1} to the previous PoA_i; this informs about HAR neighbourhood information in the reverse direction, with respect to the MN's movement. Annex E.3 provides a detailed description of both direct and indirect HAR updates together with minor optimisations to boost HARD state convergence.

Considering the number of moving MN's across different PoAs, it can be seen that the above HAR discovery steps provide fast convergence of HARD state across all members of the joint R-M-neighbourhood. Each PoA_i is expected to have at least one[5] MN attached to its network in transit from PoA_{i-1}.

It can also be seen, that identification of HAR neighbourhood is resolved *cooperatively* by all bypassing MNs; information hinted by an MN does not typically assist the MN itself, but *other* MNs travelling in the reverse direction. Since each MN typically travels towards a *different* destination, it encounters different PoAs. The disparity in the travel path of each MN's spread across different PoAs, allows ARs to identify their HAR neighbours fast and robustly, while subsequently informing MN's about potential HAR candidates. Annex E.4 presents a mechanism to support robust resolution of malicious HAR hints during HAR discovery.

[5] In GSM networks a Message Switching Centre (MSC) which is the equivalent of an Access Router (AR) in IP networks, can handle typical loads of in-transit MNs in the order of 2300 nodes/sec

4.7.3 The 'capability' perspective of m-routing state

It is interesting to observe that availability of m-routing state between HAR neighbours, can support also the notion of IP service *capability* pertaining to some measure of service utility essential for assessments during PoA handoff selection. ARs can make available in a metric-oriented manner *perspectives* of their network-layer surroundings, as an added-value handoff criterion effective within the local geographical vicinity of the MN.

The notion of service capability within a WISP domain may not be limited to seamless coverage. Service or performance capabilities pertaining to link or IPv6 layer may also characterise the geographical locality at which the MN is residing. The availability of m-routing state can support mobility-aware routing not only in the context of expediting the MN's IPv6 handoff, but also for *any* service or performance *capability* that can be supported by HAR neighbours. In this manner, added-value IP service enhancements for the MN, can be managed dynamically.

With the capability perspective in mind, the RNV-MNV mappings of HAR neighbour may be augmented to encompass service or performance capabilities in any context of IP connectivity. This can be achieved through attribute-value tuples applicable to some IP connectivity context of any service *ontology* [239], according to the following format:

$$\begin{aligned} MNVe_{k,i} = \{&'Link\ Ontology', ('coverage', ('AP\ Identity', APID_i), \\ &\qquad ('AP\ Location', \theta_{APIDi})), \\ &, (CTX_2, capLen_2), \\ &\quad \vdots \\ &, (CTX_q, capLen_q)\} \end{aligned} \tag{4.5}$$

$$\begin{aligned} RNVe_{k,i} = \{&('Network\ Ontology', ('roaming', ('AR\ IPaddress', IPaddr_{k,i}) \\ &\qquad ('AR\ PrefLength', PLen_{k,i}), \\ &\qquad ('AR\ L2addr', LLA_{k,i})), \\ &, (CTX_2, capLen_2), \\ &\quad \vdots \\ &, (CTX_q, capLen_q)\} \end{aligned} \tag{4.6}$$

$$where \quad (CTX_q, capLen_q) = (CTXtype_q, capabilty_{q,1}, capability_{q,2}, \ldots, capability_{q,n}) \tag{4.7}$$

$(CTX, capLen)_q$ is the tuple containing the context type, $CTXtype_q$, together with a number of supported capabilities $capability_{q,n}$ for that context.

The notion of IP service capability can be exploited in a mobility-aware IPv6 domain, both by the network and the MN; they can utilise it to support *choice* between available forms of IP service provisioning ahead of the MN's upcoming IPv6 handoff.

As a result, choice between IP service capabilities can be embedded in MN's handoff decision. Such form of handoff manipulation can be subsequently influenced by selection criteria or policy rules [227] representing user preference or network dynamics. In this manner, choice and its enforcement mechanisms can establish the basis of *hindsight* towards advance forms of intelligent IPv6 handoffs.

The need to manipulate or act over the IP service capability space calls for the definition of a capability-based language specification for mobility-aware IP services. The above augmented MVN/RNV-element syntax lays some basic foundations towards such a specification. An in-depth capability syntax specification is beyond the scope of this thesis. Nonetheless, it is an important direction for future research, towards advanced forms of intelligent proactive handoff management that merits further investigation.

4.7.4 Service disparity during an IPv6 handoff

Current mobile systems support handoffs based predominantly on reactive physical-layer properties such as signal-to-noise ratio (SNR), bit-error-rate (BER), or received signal strength (RSS) [240]. We argue that *in future ubiquitous wireless Internet, the MN will need to consider the capabilities of each point of IP attachment in relation to the user's requirements, as an integral part of any handoff decision.* This is certain to have a much stronger impact where multi-WISP network deployment is overlaid within the same geographical area. Such deployment can be manifested either as homogeneous or heterogeneous networks.

It is important to note that *heterogeneity* is not limited to integration of different access technologies such as GPRS, 802.11x, WiMax or Bluetooth. Heterogeneity is also manifest above the network layer. For instance, competing metropolitan WLAN vendors (e.g WiMAX, 802.16) [241] may offer short-term special tariffs for raw connectivity to manage reduced system utilisation of their domains.

Alternatively, hot-spot WLAN ISPs (WISPs) targeting key application services such as interactive gaming in densely populated areas, may be supporting proxies at the edge of their domain to transcode multimedia streams transparently [242], enabling access to small wireless devices. Each class of mobile users, may have differentiated perspectives of service and subsequently, *handoff* utility; for some, the most important handoff criterion becomes cost, while for others sheer application-level performance. For this class of users, the best handoff AR becomes the one maximising their personal/corporate utility; this is significantly different than monolithic signal strength handoff assessment.

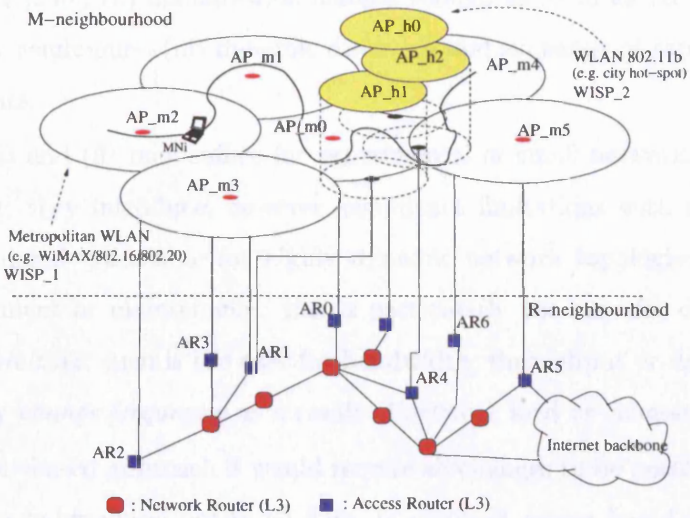


Figure 4.6: Competing WISPs offering WLAN service over the same geographical area

The key towards embracing such form of heterogeneity, through network-level handoff diversity in an mobility-aware provisioning scenario, is to provide the MN with enough information concerning IP and/or application-level services from competing WLAN ISPs, so that it can make informed handoff decisions based on its own selection criteria. This is illustrated in figure 4.6; competing WLAN ISPs offer wireless technology hybrids, provisioning long-range 802.20 wireless metropolitan-area networks (MAN), or short range 802.11a/b/g hot-spots installed in densely populated areas. Each WLAN provider supports its own set of service capabilities in contexts such as tariffs, emergency services [243], applications, quality, or security over a unified WLAN medium.

Abandoning the traditional fixed long-term subscription approach, each network *advertises* its capabilities directly to the MN device, through proactive capability context establishment. In a manner similar to 'pay-as-you-go' or 'pay-per-view' schemes dynamic subscription schemes [244], the MN is set free to choose the preferred WISP for its next IPv6 handoff, based on its own private or corporate selection policy.

To achieve this goal, there are two primary alternatives: (i) allow the MN to query each AR reactively as it associates with the respective PoA, or (ii) push *proactively* capability-information valid within the local HAR vicinity, towards the MN through its current AR. Under either of these approaches, a reactive IPv6 handoff meets severe service disruption as well as scalability problems.

For the proactive approach, discovery of capabilities amongst HARs may be supported through: (i) centralised mapping and capability referral through a domain wide back-end server [245], (ii) domain-wide manual configuration of all AR with the capabilities of their neighbours, (iii) dynamic discovery and exchange of capabilities among HAR neighbours.

Options (i) and (ii) may suffice for experimental or small networks operated by a single provider; they introduce, however, significant limitations with respect to their operation, rendered unscalable for highly dynamic network topologies as a result of failure, deployment or maintenance; this is particularly the case for capabilities that exhibit *high volatility*; such is the case for bandwidth, throughput or delay capabilities which typically *change frequently* as a result of network load or congestion.

For a server-based approach it would require all changes to be posted to the central server and then re-broadcast out to all ARs. In addition, server-based schemes present a single point of failure for the protocol. For domain-wide configuration highly-volatile capabilities become impractical to manage through centralised servers[6].

On the contrary, dynamic HAR discovery supported through m-routing state manipulations, offers a flexible capability exchange mechanism between cooperating WISPs, extending current PCS cellular roaming models. Cellular providers allow subscribers to roam among competing networks sustaining continuous service. Although network selection may be automatic, the selection (i.e. handoff) criterion remains associated solely to physical layer properties as opposed to other criteria such as cost efficiency. From a user perspective, this is particularly important in modern provisioning environments, since *most overlaid (competing) cellular networks provide near-similar quality of signal strength.*

Thus far, we have identified the means by which handoff ARs establish m-routing

[6]growing referral signalling overheads which must be counterbalanced with the validity of the capability metric.

adjacency for MN mobility management purposes. It is now important to identify how ARs 'conspire' to establish or relocate IP connectivity state for the roaming MN proactively, ahead of its next IPv6 handoff.

In the following section, we present the mechanism through which establishment of state, pertaining to the IP connectivity of the MN, is effected between HAR neighbours. Such state describes only one capability context: that of *addressing and routing*, for the purposes of *seamless* IPv6 transition between visited networks within the current R-neighbourhood. As described previously, the type of state supported may be augmented to include *any* kind of capability context that is available at the next handoff AR and of potential benefit to the MN's utility function during its next IPv6 handoff.

4.8 Proactive Context-state establishment

To support, seamlessness during MN's IPv6 handoff, proactive IPv6 handoff management requires the discovery of HAR neighbours, one mobility-hop away from MN's current PoA. This is achieved, as described previously, by identifying the M-neighbourhood and its respective virtual R-neighbourhood mapping; the handoff management function identifies the complete list of ARs that emerge as IPv6 handoff candidates for MN's next IP cell transition.

Determination of MN's current HAR neighbourhood is followed by forward identification of state information pertinent to MN's context of IP connectivity for its next IP handoff; state identification is pursued by the current AR (AR_c) for any capability context where seamlessness is *requested* by the MN. It is intuitive that if seamlessness is not a required capability, then reactive MIPv6 handoff management is sufficient for MN's operations.

Once one or more contexts of IP connectivity state have been specified by the MN to AR_c, the request is brokered to all (or some) members of the R-neighbourhood, ahead of the MN's next IPv6 handoff.

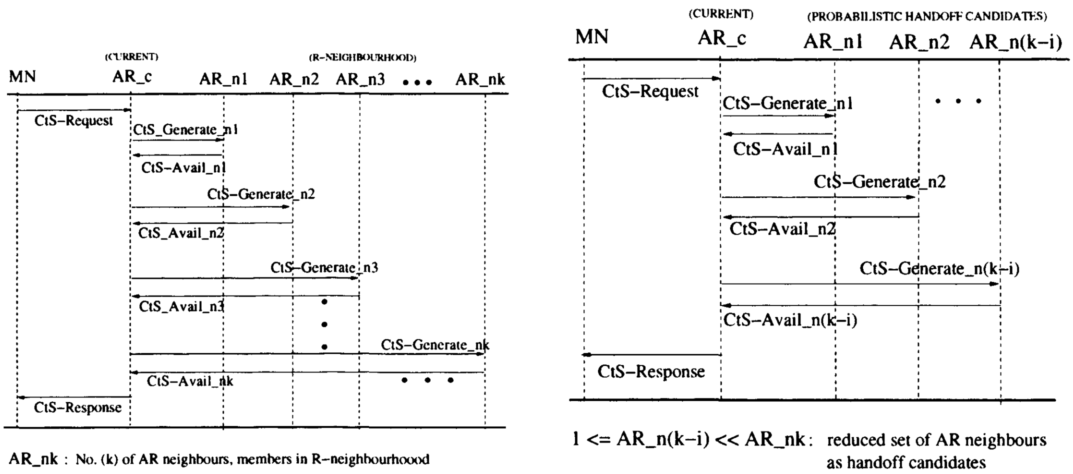
This section presents the means by which members of the R-neighbourhood *conspire proactively* to establish the required IP connectivity state essential to the MN's IPv6 handoff in support of delay seamlessness capabilities.

4.8.1 Generic context state establishment mechanism

Before dealing with context-specific instances of IP connectivity state, we present first the abstract state establishment mechanism exploiting the availability of the HAR neighbourhood. The generic state establishment mechanism is then employed to configure proactively MN's IP Roaming state, demonstrating, thus, its application in any IP connectivity context.

Upon completion of the MN's handoff, an AR transits from state new (AR_n), to current (AR_c). At this point, the MN first *identifies* itself to AR_c specifying the *type* of (capability) context required towards its next IPv6 handoff; this is presented to AR_c through a *CtS-Request* message. Depending on the type of context, the MN provides a context initialiser to AR_c specifying the requirement for state establishment or relocation. Relocation is relevant in state contexts that do not require necessarily re-establishment; such is the case for instance, with header compression state [246].

The current AR (AR_c) transmits[7] the context initialiser of the MN to all neighbouring HARs in a *CtS-Generate* message. HAR neighbours receive the state generation request and depending on the type of requested context, apply the relevant state establishment mechanisms; for instance AAA context type employs a different state establishment mechanism than QoS or robust header compression [247]. In this manner, context state of interest to the MN is established through HAR neighbours in preparation of MN's next IPv6 handoff.



(a) State Establishment to entire R-neighbourhood

(b) State Establishment to candidate AR set only

Figure 4.7: State establishment for entire R-neighbourhood versus minimal candidate AR set of the R-neighbourhood

Once the state has been established, each of the AR neighbours responds back to AR_c with state information and a *handle* for the type of context state estab-

[7]the transmission may be effected either through unicast or multicast communication. From a seamlessness perspective there is no significant gains other than packet duplication for the CtS-Generate

lished/relocated; this is done through a *CtS-Avail* message. A handle provides proof that state has been established at an AR neighbour and awaits activation upon the handle's presentation to the AR.

All state establishment responses from HAR neighbours are collated by AR_c into a HAR neighbourhood *state tuple* supporting the following generic format:

$$\begin{aligned} CtS_{MN_n} = \{ & (CtState_1, AR_1, APID_1, Channel_{APID1}), \\ & (CtState_2, AR_2, APID_2, Channel_{APID2}), \\ & \qquad \vdots \\ & (CtState_i, AR_i, APID_i, Channel_{APIDi})\} \end{aligned} \tag{4.8}$$

for each type of context requested by the MN. $CtState_i$ is the state generated by AR_i controlling the coverage area $APID_i$, which in turn operates at frequency $Channel_{APIDi}$ within the respective M-neighbourhood mapping. The mapping between AR_i (sender of $CtState_i$) and $APID_i$, can be quickly identified by the TMM table maintained by AR_c as a result of HAR discovery. The *state response tuple* generated at AR_c for the particular context, is subsequently pushed to the MN through a *CtS-Response* message. Annex E.5 presents a brief analysis on the measure of proactivity in signalling deliberations in the proposed mobility management architecture.

4.8.2 Context state establishment paradigm: IP Roaming state

Addressing and routing, identified collectively as *IP Roaming state*, comprise the fundamental type of MN's IP connectivity context onto an IPv6 wireless network; without this the MN remains disconnected from the network while its IP flows are severely disrupted. Instead of establishing such state reactively during MN's next IPv6 handoff, the proposed handoff management model pursues its proactive establishment over the R-neighbourhood reachable from MN's AR_c.

Assuming existence of home address and bindings to some HA, the MN issues a *CtS-Request* to the AR_c specifying 'IP-roaming' as the required seamlessness context; the state request includes also its *link-layer address (LLA)* as the context initialiser.

On receipt of the CtS-Request, AR_c sets out to *broker* the particular type of context to its HAR neighbours; this is done by forwarding MN's context initialiser (MN's LLA) to each of its RNV members through a *CtS-Generate* message.

In the context of IP-roaming, the state requested by AR_c (on behalf of the MN), comprises of a set of unique IPv6 unicast care-of addresses (CoA) and their respective default routes, for the MN to be admitted.

Each of the HAR neighbours receiving the CtS-Generate request, is empowered with the IP-Roaming state generation task (for the MN), in a *proxy-stateless* manner (see Annex E.6.1). Such task comprises of 3 distinct functions, essential for reachability under IPv6:

- globally-reachable IPv6 CoA generation for the purposes of global-reachability.
- link-local IPv6 CoA generation for the purposes of both address resolution and (link-local) neighbour reachability.
- Adjustment of Neighbour Cache to support address resolution, neighbour reachability and proxy-neighbour discovery [175] of the globally-reachable IPv6 CoA (to be) configured for the MN.

For global IPv6 CoA generation, the HAR neighbour combines the link-layer address (LLA) [110] of the MN into an IPv6 *soft CoA (sCoA)*; following generation of the sCoA, the HAR neighbour performs duplicate address detection on that address [110] through a standard neighbour solicitation on it link.

For link-local CoA generation, the HAR neighbour combines the LLA of the MN into an IPv6 *soft link-local CoA*. Such address, together with MN's LLA are used to adjust proactively the Neighbour Cache of the AR neighbour as prescribed by standard IPv6 Neighbour Discovery. The AR neighbour adjusts its cache with a *soft neighbour-cache entry* containing the link-local IPv6 CoA and LLA address of the MN, which is flagged as as *proactively reachable* or `P-REACHABLE`. Subsequently, the AR neighbour 'defends' this IPv6 against the sCoA against duplicates until the MN handoffs on its link or the sCoA lifetime expires.

It is intuitive that, since DAD for proxy-stateless generated sCoA is performed by the AR neighbour in advance of MN's IPv6 handoff, the MN will not experience DAD latency (1000ms) *during* its next IPv6 handoff. Hence, the proposed mechanism is guaranteed, by-design, to eliminate DAD delay from MN's IPv6 handoff process.

Once the aforementioned 3 functions have completed, each HAR neighbour, prepares a state tuple, describing itself as candidate point of attachment (PoA). This is identified as *PoA tuple*, and takes the form:

$$CtS_{AR_i} = ('Roaming', sCoA_i, DefaultRoute_i, APID_i, Channel_{APIDi}) \tag{4.9}$$

where '*Roaming*' identifies the type of capability context, $sCoA_i$ is the soft IPv6 CoA generated at the HAR neighbour; $DefaultRoute_i$ is the route applicable to that $sCoA_i$, whereas $APID_i$ is the identity of the AP corresponding to that HAR neighbour operating at frequency $Channel_{APIDi}$. The above tuple is sent to AR_c originating the CtS-Request, through a *CtS-Avail* message.

All state establishment responses from HAR neighbours are collated by AR_c into a *IP roaming-state tuple*, comprising of the following generic format:

$$\begin{aligned} CtS_{MN_n} = \{ \ & CtS_{HAR_1}, \\ & CtS_{HAR_2}, \\ & \quad \vdots \\ & CtS_{HAR_i}, < FlowForwardingState >, RNVe_c\} \end{aligned} \tag{4.10}$$

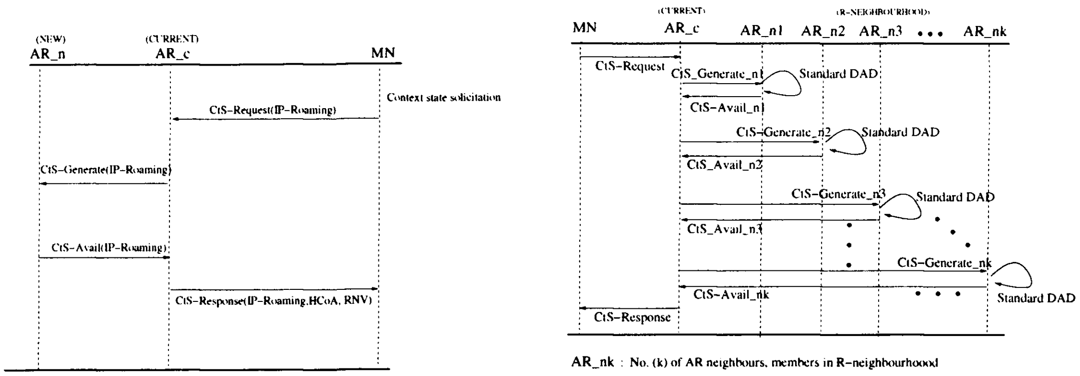
CtS_{HAR_i} represents the state tuple generated by the i-th HAR neighbour supporting the following format:

$$CtS_{HAR_i} = ('Roaming', sCoA_i, DefaultRoute_i, APID_i, Channel_{APIDi}) \tag{4.11}$$

CtS_{MN_n} comprises MN's collective IP-Roaming state, which is subsequently sent to the MN via a *CtS-Response* message; it encompasses also flow forwarding state generated by AR_c as well as its own RNV element. Flow forwarding management deals with sustaining seamless packet flow towards the MN and is presented in detail in Chapter 5. Figure 4.8(a) illustrates the signalling exchange for the purposes of IP Roaming establishment.

The MN stores the received RNV-element for the purposes of Handoff AR discovery, as described in Algorithms 1 and 2 of Section E.3.1. In the event that Algorithm 2 is followed, AR_c does not provide its own RNV-element.

Annex E.6 presents operations intrinsic to proactive IP Roaming state establish-



(a) IP-Roaming state establishment (b) Standard DAD over proactive signalling

Figure 4.8: IP Roaming state establishment and the encompassing of the standard DAD process over proactive signalling, in advance of MN's IPv6 handoff.

ment, such as Duplicate Address Detection (DAD) and state collection at MN's serving AR.

4.8.3 Coupling established state with handoff management

State established proactively must be tightly bound to the critical process of IPv6 handoff between two neighbouring PoAs. To achieve this, it is first essential to acknowledge that a link-layer (L2) handoff between the respective APs in the M-neighbourhood, occurs *before* the IPv6 handoff can be initiated.

Determination of a mapping between the identity of the serving link-layer device such as an AP and the respective handoff AR serving at the network layer, allows the immediate derivation of one from the other, as soon as one of the two identifiers becomes available within a point of attachment. Since an L2-handoff occurs before an IP handoff, a host may deduce instantly handoff state about the network layer, namely the AR, as soon as its underlying link-layer (AP) has been identified.

Proactive IPv6 handoff management supports *by design* such mapping within the established joint R-M-neighbourhood of an MN, for all immediate[8] candidate handoff AR neighbours. From the pre-established IP-Roaming state (see Section 4.8), there exists a unique such mapping between the handoff AR candidate and the identifier of the corresponding link-layer segment (AP candidate), comprising collectively a single PoA (see Section 4.6.1).

This mapping may be utilised by determining [9] first the identity of the associating

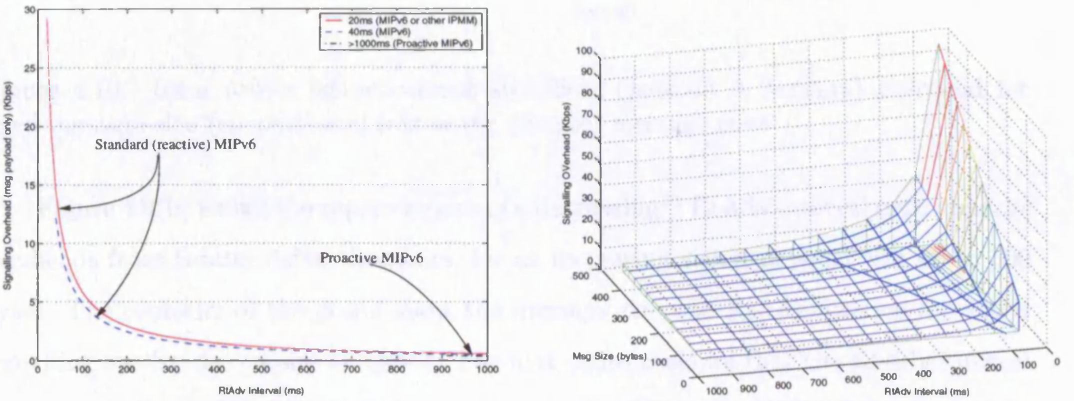
[8]i.e. in the vicinity of MN's next IPv6 handoff

[9]Determination of the BSSID is readily available as L2-management APIs for most popular operating systems such as Windows and Linux [248]

AP during MN's L2-handoff at the *new* PoA. In 802.11 WLAN networks such identity is found through the Basic Service Set identifier (BSSID). In cellular networks such identifier is available through the cell global identity[10] (CGI).

Combined with its pre-allocated IP-Roaming state, the AP identifier enables the MN to select instantly the *correct* PoA tuple from its IP Roaming state, for the purposes of its subsequent L3-handoff at that PoA. A PoA tuple contains the soft CoA, the default route, and Neighbour (i.e. link-local) reachability state enabling the network-level attachment at that particular PoA.

It is important to emphasise that *immediate activation of L3-handoff state is effected with no dependence of network layer signalling on the part of AR*; that is, *with no* reliance on router solicitations/advertisements, traditionally used for the purposes of IPv6 movement detection. In fact, under proactive IPv6 mobility management, the movement detection function may be eliminated, together with overheads arising from the increase of the router advertisement rate, prescribed by MIPv6.



(a) Current RtAdvert Msg Size (73 bytes)

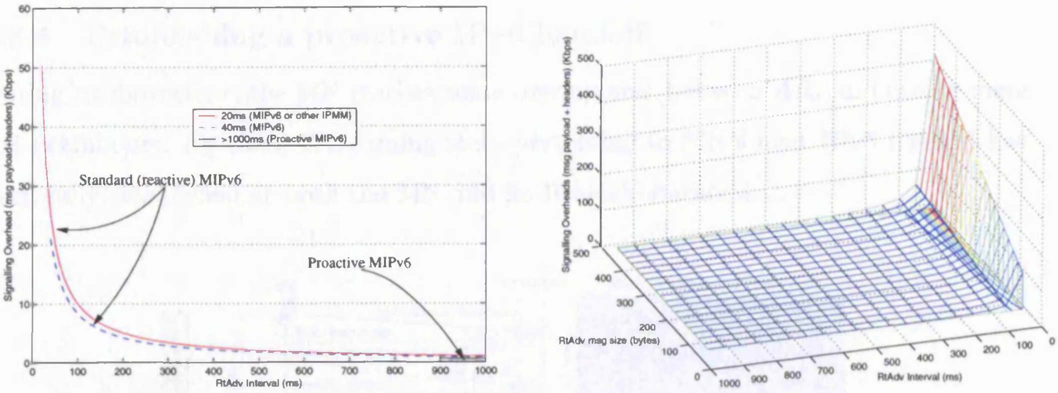
(b) Increasing RtAdvert Msg Size (up to 500 bytes)

Figure 4.9: Router Advertisement signalling (*payload only*) overhead for fixed message size (current) and increasing (future) message sizes

Figure 4.9(a) presents the difference in signalling payload overheads between the MIPv6 standard, or any scheme that requires further reductions of the Router advertisement interval, and Handoff Management under Proactive Mobile IPv6. It can be seen that MIPv6 requires a minimum of 14 Kbps of the wireless link only for transmission of periodic RtAdvertisments. Any scheme (including MIPv6) that attempts to speed up

[10] CGI comprizes of Cell identity (CI) and Location Area Identity (LAI) or Routing Area Identity (RAI) for the purposes of paging and routing.

Movement Detection by reducing the Router Advertisement interval down to 20ms will double signalling overheads up to 30 Kbps. Proactive IPv6 handoff management eliminates such requirement by pushing the router Advertisement interval back to values $\geq 1000ms$ and a reduction of RtAdv signalling overhead by two orders of magnitude (down to 0.584 Kbps).



(a) Current RtAdvert Msg Size (73 bytes)

(b) Increasing RtAdvert Msg Size (up to 500 bytes)

Figure 4.10: Total router advertisement signalling (*payload + headers*) overhead for fixed message size (current) and increasing (future) message sizes

Figure 4.9(b) shows the repercussions of a decreasing[11] RtAdv interval onto message overheads from Router Advertisements, for an increasing signal payload size up to 500 bytes. The contours of the graph show the message size and RtAdv interval for which signalling overheads remains constant. The first contour shows that the RtAdv interval must vary between 100-400ms, with a message payload between 73-500 bytes to sustain a constant measure of signalling overheads around 7 Kbps.

Figure 4.10(a) shows the total signalling overhead (message payload plus packet headers) incurred by the same classes of router advertisement intervals shown in figure 4.9(a). It can be seen that headers contribute significantly as the router advertisement interval decreases; this is as much as 7 Kbps for MIPv6, but only 0.42 Kbps under Proactive IPv6 mobility. In the event that the RtAdv interval halves down to 20ms the header induced overhead increases up to 20 Kbps.

Figure 4.10(b) shows that total signalling overheads increase by almost a factor of

[11] The Mobile IPv6 working group in an effort to enhance movement detection has been constantly reducing the RtAdv interval. While the MIPv6 standard specifies currently an average RtAdv interval of 40-50 ms, its is possible that future versions of the MIPv6 specification (or other mobility management mechanisms) may attempt to specify a lower RtAdv interval for that purpose.

4, as the size of RtAdv message increases (up to 500 bytes for the entire packet). It becomes clear that *proactivity in state establishment, coupled tightly with the handoff management function, provides significant reductions in signalling overheads incurred by router advertisements alone. In fact, proactive IPv6 handoff management removes reliance on core IPv6 Neighbour Discovery eliminating altogether the need for movement detection.*

4.8.4 Establishing a proactive IPv6 handoff

During its movement, the MN reaches some overlap area between AR_c and one or more AR neighbours. By then, IP roaming state pertaining to MN's next IPv6 handoff has been fully established at both the MN and its R-neighbourhood.

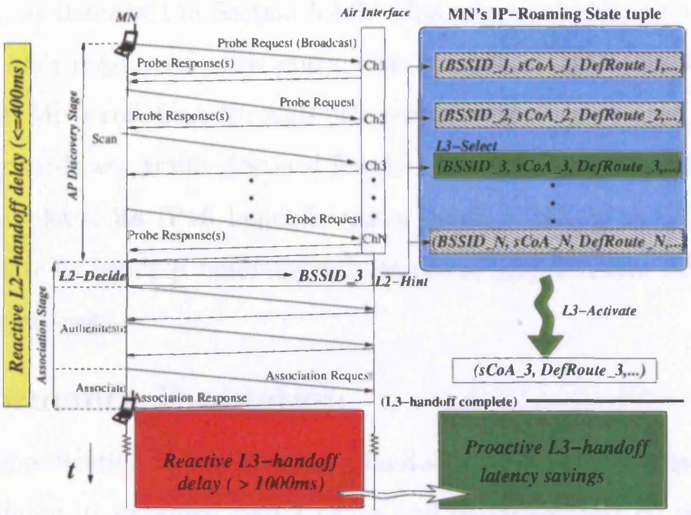


Figure 4.11: Integration of L2-handoff hint during proactive IPv6 handoff management

Upon exceeding the range boundaries of its serving coverage area manifested as poor SNR, the MN deassociates with its current AP, initiating a link-layer handoff. During MN's link-layer handoff the network-layer handoff management function is provided with the identity of the AP selected for association in the form of a link-layer hint (*L2-hint*).

An example of such an L2-hint is shown schematically in figure 4.11; during a link-layer handoff, the MN engages into a AP discovery phase where channels are successively scanned for AP availability. Past the AP discovery phase the MN selects the AP to associate with, based on the measure of the AP's received signal strength/quality. While the MN decides which is the AP to be associated with (`L2-decide` period), an L2-hint ($BSSID_3$) is fed to the proactive IPv6 handoff management function, informing about its AP identity. The L3-handoff function matches instantly the AP identifier with the respective PoA tuple configured as part of the IP Roaming state of the MN. Subsequently the correct PoA tuple is selected (`L3-Select`) and activated (`L3-Activate`), by means of the respective soft CoA, set as its current *primary* care-of address.

It can be seen that the delay period (red) incurred during an L3 handoff under standard Mobile IPv6 can be successfully eliminated under Proactive IPv6 mobility management (green); this comes in effect with no dependence on neighbour discovery signalling emanating from the candidate AR.

The MN attains link-local reachability by utilising the (re)-association trigger at the new PoA. In particular, as soon as the AP has registered a new association for the incoming MN, the AP initiates an `ATTACH` L2 trigger to the connecting interface of the controlling AR, as described in Section 5.4.2. This trigger at AR_n removes the P flag from the new AR's neighbour cache entry; this is set now to `REACHABLE`, while AR_n ceases to defend MN's soft CoA through proxy-neighbour discovery; at the same time, the new AR forwards any traffic destined for the MN, over the local link.

Having completed its IPv6 handoff with a single P-NA to AR_n, the MN then proceed to update its location bindings with its peers, both HA and CN(s), through a binding update message.

4.9 Performance Evaluation

Following the specification of the proactive handoff management architecture, we now proceed to evaluate its expected performance and hence support or reject the initial hypothesis on two individual accounts as stated in Section 4.2.1:

- *proactivity in handoff management can address effectively transmission delay seamlessness during MN's IPv6 handoff with respect to the delay requirements of MN's interactive multimedia communication with its peers.*

- *proactivity in handoff management can increase efficiently MN's operational utility as a measure of received service expressed through MN's own performance criteria or policies.*

With respect to delay seamlessness, the performance analysis investigates aspects of non-determinism in the mobility pattern of the MN and its impact on handoff performance, as well as convergence in handoff AR discovery (HARD). These are expressed as a varying measure of MN *speed* and *pause period* in its movement pattern. In addition, handoff performance and HARD convergence is investigated for varying MN densities within a geographical region.

With respect to handoff selectivity and its impact on the measure of MN's handoff utility, we investigate the influence of the size of the PoA topology within a geographical area. This is because handoff utility is expected to vary when there exists considerable diversity in the choice of PoA among the set of candidate handoff PoAs. With regard to service utility and handoff selectivity it is important to differentiate between *handoff neighbours* and *handoff candidates*; handoff neighbours are the set of neighbouring PoAs that can potentially accommodate the MN in its next IPv6 handoff, depending on MN's movement direction. On the contrary, handoff candidates are the set of PoAs whereby, the MN falls immediately within their transmission range *during* its next IPv6 handoff. It follows that handoff candidates are necessarily handoff neighbours. By contrast, handoff neighbourhood is not a sufficient condition to justify handoff candidacy; it requires further that the MN is also found within its transmission range during its next handoff, to justify handoff candidacy.

4.9.1 Methodology

To assess the performance of proactive handoff management we pursue a protocol performance analysis by means of discrete event simulations [249]. Subject to satisfactory performance, a real-kernel implementation of the proposed architecture is the objective of future work.

For the purposes of simulations, the mobility model of NS-2 [250] was extended first to encompass key parts of proactive handoff management as well as standard IPv6 mobility management for the purposes of comparison. These parts implemented in particular:

- indirect and direct RNV updates; these messages are effected between the MN and the AR (indirect) and between ARs (direct) respectively.

- proxy-stateless generation of multiple[12] Care of Addresses (CoAs) per mobile node.

- handoff initiation as a result of the MN moving out of the range of its current PoA and proactive handoff control by means of mapping L2 AP identifiers onto the correct CoA matching the subnet prefix of the respective AR neighbour.

[12] Typically, NS-2 allows only a single, static address to be assigned to each node

Core IPv6 protocol functions pertaining to router solicitation/advertisement from Neighbour Discovery (ND) protocol were implemented for the purposes of standard MIPv6 operations.

4.9.2 Network topology (Routing Neighbourhood) model

Determining a correct network topology model that represents the Internet with reasonable accuracy is essential for the correctness of derived simulation results. To identify a realistic network topology model for the purposes of simulation, we turn our attention on synthetic network topology models [251] that track the behaviour and node dynamics of Internet [252] emerging from real network topology measurements.

Measurement of Internet topology are mostly done using BGP tables or by means of ICMP traceroutes [253]. BGP tables are routing tables employed to effect routing between Autonomous Systems (AS) with each AS typically comprising one or more provisioning domains. In most cases, however, BGP tables are private and hence the topological view of the Internet remains restricted. This is also the case for *route tracing*, since not all routers enable responses to ICMP messages. By means of analysis of local routing regions, Shenker et al [254, 255] report that at least 25% to 50% of the links are not encompassed in current Internet topology measurements; their results imply that the attained view of Internet topology by such measurements is at best incomplete.

Nonetheless, focusing on the node degree of the network topology, Broido and Claffy [256] report from a set of 20 participating traceroute and BGP table monitors a network topology of 665,000 nodes with an average node degree[13] of about 4. Faloutsos et al [257] report that the distribution of the node degree of the ASs in the Internet follows a power law of the form $y = x^{\alpha}$, with exponent -2.2, implying that the number of nodes with degree d is approximately $1/d^{2.2}$.

Such power laws have been used to validate the accuracy of a given algorithms in generating representative Internet topologies. To this end, Barabasi and Albert [258] propose important adjustments to network topology generation, by introducing two important factors that appear to justify the existence of power law relationships in any network topology: *incremental growth* and *preferential connectivity.* Incremental growth refers to open networks that form through continuous addition of new nodes, increasing, thus, gradually the size of the network. Preferential connectivity refers to the tendency of a new node to connect to existing nodes that are either highly connected or popular.

[13]The degree of a node in a graph represents the number of its neighbours.

Medina et al [259] in their experimental investigation demonstrate that the synthetic network topology generation models such as the Waxman [260] and the transit-stub models that have been typically available through tools like the GeorgiaTech topology generator (GT-ITM) [261], TIERS [262] or Inet [263, 264] do not appear to be representative of Internet topologies as they exhibit low correlation (< 0.85) with respect to power laws identified (correlation > 0.95) by Faloutsos et al.

The authors augment further the importance of Barabasi and Albert findings, by reporting that synthetic topologies generated *without* preferential connectivity nor incremental growth exhibit low correlation with the measured topologies exhibiting power law relationships. They show that, preferential connectivity appears to be a necessary condition for the power law relationships to hold; the presence of incremental growth increases the correlation coefficients. Furthermore, both preferential connectivity and incremental growth are required for the out-degree power law to exist.

Node assignment onto the topology plane exhibits more realistic Internet topologies for heavily-tailed[14] distribution, implemented through a bounded Pareto distribution:

$$f(n) = \frac{\alpha k^{\alpha} n^{-a-1}}{1 - (k/P)^{\alpha}} \tag{4.12}$$

Configuration of the aforementioned parameters and distributions is encompassed in the software implementation of the Medina et al optimisations, onto existing Waxman and Albert & Barabasi synthetic topology models [265]; this implementation effort is identified as Boston topology generator (BRITE) [266]. The output network topologies produced by BRITE, exhibit the essential power-law relationships described by Faloutsos et al, while supporting incremental growth as well as preferential connectivity. *The aforementioned topology generation rationale and the associated metric features justify the use of BRITE for the purposes of network topology generation in subsequent performance evaluation of proactive IPv6 mobility management through simulations.*

The topologies generated are identified as synthetic given that they comprise of two tiers: definition of Autonomous Systems (ASs) and definition of Router topologies within each AS. For the purposes of simulation the number of ASs is chosen to be small, since we are interested on round trip time delay within each AS as opposed to the complexity of the meshing between ASs. The AS topology is generated by means of a Waxman topology model whereas the router topology embraces the Albert and Barabasi

[14] in the case of synthetic network topology generation, a heavily-tailed distribution represents AR clusters onto the assignment plane, as opposed to random placement across the entire plane.

(A&B) topology model. The A&B topology model in the inter-AS router topology is chosen as a result of better node degree distribution amongst the connected ASs [267, 268]. Given that the number of ASs is small, such a requirement is not essential for the generation of the AS topology tier. Table 4.2 present the set of parameters employed for generation of all simulated topologies. The sole parameter of variance in these topology models is the number of AR nodes. Parameters α and β represent the Euclidian distance between two nodes in the Waxman probability model of connectedness between two AR nodes [265, 260].

AS Topology Parameter	Value	Router Topology Param	Value
High grid Sqr (HS)	[100,1000]	HS	[5,10]
High grid subSqr (LS)	[10,100]	LS	[2,5]
N (no of nodes)	[10,1000]	N	5
AS topology Model	Waxman	Router topology Model	Albert&Barabasi
Node Placement	Heavy-tailed	Node Placement	Random
Grow Type	Incremental	Growth Type	Incremental
Preferen. Connect	yes	Preferen. Connect	yes
Bandwidth Distr	Constant	Bandwidth Distr	Exponential
Min Max B/w	[1,10] Mbps	Min Max B/w	100 Mbps
Edge Connect	Random	Edge Connect	Smallest Deg.
alpha	0.15	alpha	N/A for A&B
beta	0.2	beta	N/A for A&B
m (no. of links/new node)	[1,4]	m (no. of links/new node)	[1,4]

Table 4.2: Synthetic topology generation parameters

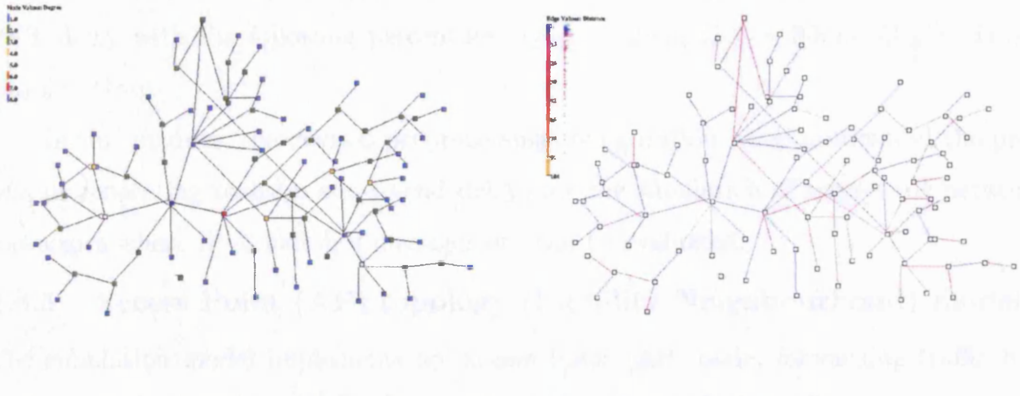
Figures 4.12(a) and 4.12(b) illustrate the maximum node degree and the (colour-coded) distance distribution, for the network synthesised, simulating a topology of 89 ARs, with a leaf (1-degree) AR set of 45 PoAs.

The distance distribution 4.12(b) within the generated network topology model is subsequently coupled with a delay distribution (see Section 2.3.3), representing the one-way delay component over the path between the corresponding and the mobile node for the purposes of subsequent packet communications. In this manner round trip times are also synthetically produced.

The mapping between topology path segments and one-way delay components is effected by means of automated preprocessing[15] prior to simulation execution. First the roots of the topology are identified[16]; these are the nodes with the highest node degree. Then all nodes connected under a root determine their distance d from the root by employing the Dijkstra shorted path tree (SPT) algorithm [269]. The distance

[15] through Matlab.

[16] in essence the roots represent the BGP routers.



(a) Maximum Node degree (b) Node (AR) distance distribution

Figure 4.12: Node degree and random node distance attributed to one way delay components

d_i is then normalised according to the maximum distance observed for that AS root. The derived factor (w_i), weighs the one-way delay value generated for that path and is applicable to the $i-th$ segment of that path.

$$w_i = \frac{d_i}{max(d_0, \ldots, d_i)} \tag{4.13}$$

The process is repeated iteratively for each segment of the path all the way to the leaf AR node (AR_k), to derive the respective delay component n_k

$$n_k = T_{gamma} w_k - \sum_{i=1}^{k} n_{i-1} \tag{4.14}$$

In this manner, the sum $\sum_{i=1}^{k} n_k$ of the delay components for each edge on the path between the root and the leaf nodes, produce the total (original) one-way delay value generated.

Intra-domain (leaf-to-leaf) delay is generated by means of a shifted gamma distribution with a maximum delay of 47-50ms. To achieve realistic inter-domain end-to-end delays, in accordance to the reported results of Section 2.3.3 and in line with observed RTT times as elaborated in Section 2.3.3 we consider maximum (heavily-tailed) end-to-end delay of 40-100ms and maximum RTT times of 80-200ms. To achieve these values we empirically adjust the location, scale and shape of the gamma distribution to values of ($\mu = 2$), ($\beta = 1.1$) and ($\gamma = 0.8$), respectively. The generated delay values are shifted positively by (a minimum of) 10ms. Such configuration produces a heavily-tailed gamma distribution with the following end-to-end delay percentiles: $Q_{25} = 25ms$,

$Q_{50} = 30ms$, $Q_{75} = 34ms$, $Q_{99.9} = 47ms$. This gives rise to an average (asymmetric) RTT delay with the following percentiles: $Q_{25} = 32ms$, $Q_{50} = 35ms$, $Q_{75} = 41ms$, $Q_{99.9} = 87ms$.

In this manner, the devised preprocessing configuration setup automated the process of generating realistic end-to-end delays for the simulation of large-scale network topologies where IPv6 mobility management can be evaluated.

4.9.3 Access Point (AP) topology (Mobility Neighbourhood) model

The simulation model implements an Access Point (AP) node, forwarding traffic between an infrastructure CN[17] and wireless MNs. Each AR from the aforementioned generated network (L3) topology as shown in Figure 4.13(a) is configured to provide IPv6 connectivity by bridging through a *single* AP over the air interface.

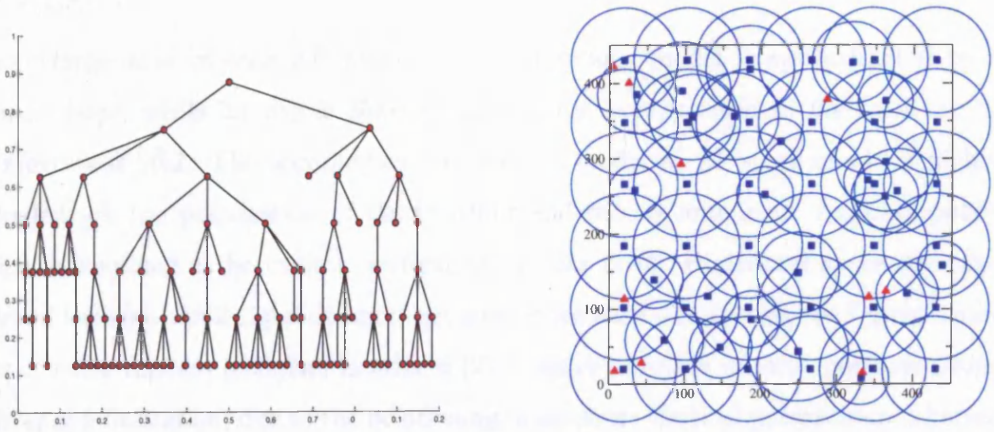
In this manner, a one-to-one mapping, between AR and AP, identifies collectively a single Point of Attachment (PoA). The set of APs enabling the devised network topology over the air interface, provide continuous spatial coverage within a 450-meter square grid. Spatial assignment of APs is performed so as to ensure: (i) minimum overlap and continuity in coverage (ii) PoA diversity by multiple APs overlapping over random grid locations (hot-spots or multi-WISP availability). The devised AP topology supports differentiation of AP transmission range on two levels:

- Basis level (n_b): a minimum number of APs are placed such that a minimum overlap of 10m is assured between APs with complete coverage over the movement grid. At the basis level all AP support a homogeneous transmission range of 70m[18].
- Hot-spot or multi-WISP availability ($n_{max} - n_b$): a number of additional APs placed randomly over the movement grid, with transmission ranges uniformly distributed between 10 and 70m

By varying the number of APs ($n > n_b$), the second level of AP topology can support the investigation of handoff selectivity as a result of PoA diversity. To this end, the mapping between ARs and APs perceived by the MN as PoA is allocated a continuous service quality metric uniformly distributed between 0 and 1. As such, simulations investigating PoA diversity during MN's IPv6 handoff can explore the relative improvement in service utility between plain-SNR reactive MIPv6 and context-aware proactive MIPv6 handoffs.

[17] placed randomly on some leaf of the network topology

[18] For a 450-meter movement grid the basis level of a simulated AP topology is 36 APs at aforementioned transmission range



(a) AR (L3) topology with AS root identified

(b) AP (L2) topology enabling L3 over the wireless interface

Figure 4.13: Perspectives of the simulated L3 (Access Router) and L2 (Access Point/Base Station) network topology Each of the leaf nodes of the network topology map onto a single coverage AP area. Red triangle represent dispersed MNs prior execution of the simulation scenario.

Each MN may associate with only one PoA at any time. Furthermore, each AP forwards data over a unique wireless channel; periodic beacons are broadcast every 100ms to MNs associated with the particular PoA, in line with 802.11 MAC sublayer specification [17].

A handoff algorithm is implemented for the transition of MNs between cells of the Mobility Neighbourhood; the algorithm is based on the modelled signal-to-noise ratio (SNR) translating into the probability of beacon reception (omnidirectional) by the MN at the particular location coordinates (on the simulation grid) away from the AP[19]. To simplify MAC interactions and simulation complexity free-space propagation (i.e. no obstructions or foliage) is assumed, since propagation delay is very small in comparison to L3 and L2 handoff delay measures.

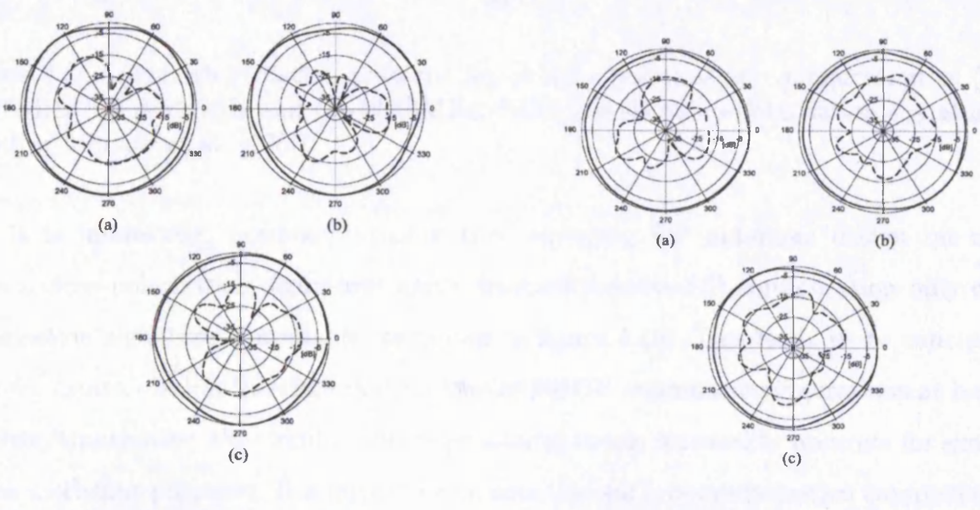
With respect to an 802.11b link-layer handoff the simulation implements the statistical delay behaviour observed experimentally in Annex D.6.1: all 13 channels are scanned with a probe request interval that is normally distributed with mean $\mu = 0.039s$ and $Std.Dev = 0.0095s$ followed by subsequent authentication and association signalling handshakes. Details of the statistical distribution of delay for the individual types of L2-handoff signals can be found in table D.2 of Annex D.6.1.

[19] normalised over the maximum SNR experienced by a wireless station over free-space propagation observed at a reference distance of 30cm from the transmitter

Cell shape

The coverage area of each AP within the AP topology model is approximated by a *circular shape*, while its *size is chosen heuristically in proportion to the total size of the movement grid.* The accuracy of the *circular* model of AP's cell shape is highly dependent on the polarisation of the transmit and receive antennae. Antenna polarisation is classified as horizontal, vertical or circular [270]. Horizontal space diversity achieved through vertical polarisation antennae have been dominating the implementations of most wireless (cellular) interfaces [271]; fewer interface implementations adopt horizontal polarisation, due to the positioning of receiver. Vertical polarisation achieves near-constant return (path) loss, while horizontal achieves varied return loss as a function of operating frequency; nonetheless, each polarisation type imposes specific antenna orientation.

Recently, circularly polarised (CP) antennae have been proposed as better[20] implementation designs, since they relieve both the AP and the MN from strict receive transmit antennae orientations [272]. At the same time, CP antennae maintain propagation loss characteristics similar to the ones of vertical polarisation and thus, appear as the natural choice in future wireless interface design [270].



(a) Horizontal polarisation

(b) Vertical polarisation

Figure 4.14: Azimuth radiation patterns for horizontal/vertical polarisation configuration at (a) 1850 MHz (b) 1920 MHz and (c) 1990 MHz. Solid line indicates co-polarisation signature while dashed line shows cross-polarisation component (Source: Morrow et al [270]).

[20] energy lost on the horizontal branch is recovered at the vertical branch

To assess the validity of the circular AP cell shape assumption we look into the type of polarisation adopted by the transmit and receive antennae; when both the receive and transmit antennae maintain the same polarisation then the radiation pattern of the co-polarisation signature becomes nearly circular over unobstructed environments. In such cases a circular model of AP's cell shape is reasonably *accurate*. By contrast, when the receive and transmit antennae have orthogonal[21] polarisation then the radiation pattern attains lobe-shaped irregularities as illustrated in figure 4.14. In such cases, a circular model of the AP's cell shape is merely a *coarse approximation* for unobstructed environments.

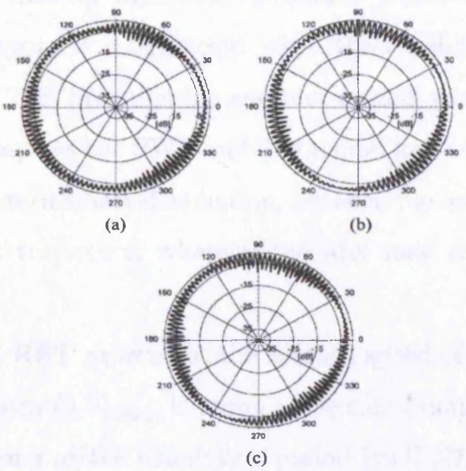


Figure 4.15: Azimuth radiation patterns for circular polarisation configuration at (a) 1850 MHz (b) 1920 MHz and (c) 1990 MHz. Solid line shows co-polarisation signature. (Source: Morrow et al. [270]).

It is interesting, however, to note that emerging CP antennae design do *not* have a cross-polarisation signature; under transmit/receive CP configuration only co-polarisation signature is available, as shown in figure 4.15. This leads us to conclude that for future wireless interface designs employing CP antennae configurations at both receiver/transmitter, the circular cell shape assumption is reasonably *accurate* for simulation modelling purposes. It is important to note that for cross-polarisation components the lobe-pattern gratings of antenna radiation pattern can be ramped up by increasing the link margin in terms of transmission power[22]. In this manner, conservative transmission range specification at peak power output for the particular interface *avoids the lobe propagation intrinsics* for omnidirectional coverage at a specific BER threshold.

[21] i.e. mixed vertical and horizontal polarisation.

[22] i.e. ramp up the transmission power to sustain the same BER

Thus, given the dominance of vertical polarisation in current antennae designs as, the emergence of CP-driven antenna designs in future wireless interfaces, as well as power control to maintain near circular coverage in any (unobstructed) direction, the circular cell shape assumption is found to be a realistic approximation for the purposes of this analysis and thus does not detract significantly from the validity of our results.

4.9.4 Mobility Model

MN's mobility pattern employs a random *way-point* (RWP) trajectory. Alternative mobility models encompass the manhattan, highway or circular terrain mobility models [273]. The choice of the RWP mobility model is guided by findings reported by Benshal et al [274]; they report that the difference in handoff probability between manhattan, random and highway layout is insignificant when the number of available changes in direction is small (1-3). This investigation assumes a small number of available changes in direction and thus, employs the RWP mobility model for the purposes of simulations.

A way-point is an intermediate destination, between the origin and final termination point of the movement trajectory, whereby the MN may change its velocity vector (speed, direction).

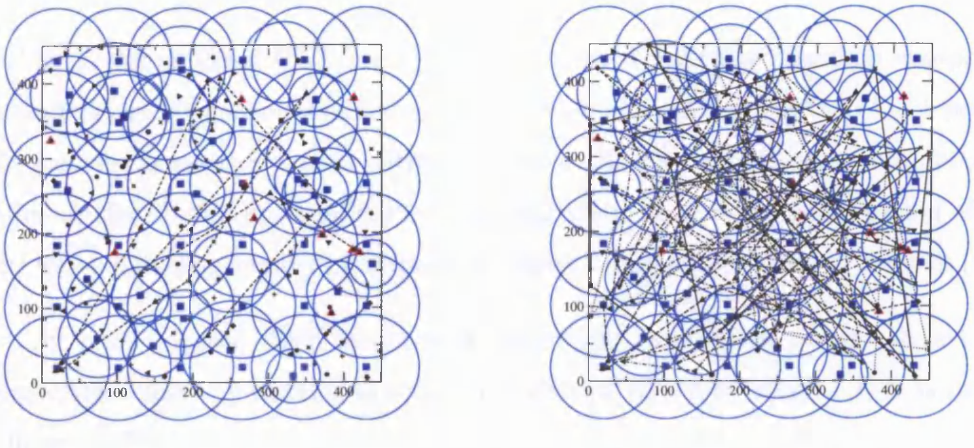
Under conventional RWP movement the average speed of MN, that is initially uniformly distributed between $(0,V_{max}]$, is found to be continuously decaying, introducing mobility transience as long as the simulation period itself [275]. Such prolonged transience affects negatively (slows-down) the rate of convergence of the proposed proactive handoff AR discovery (HARD) mechanism. MNs are observed during the course of the simulation to mobilise with steadily decaying movement speed across the topology grid, instead of maintaining a (steady-state) average speed of approximately $\frac{V_{max}}{2}$.

To this end, simulations conducted during this performance evaluation adopt the *modified* RWP model of Le Boudec and Vojnovic [276]. Their extensions the conventional RWP model address effectively the issue of MN *speed stationarity.* The modified RWP model adopted, ensures that, during proactive or reactive mobility management simulations, MNs maintain at steady-state a mean speed measure, with nearly constant variance; this is achieved by imposing a lower bound of minimum MN speed [277].

Figures 4.16(a) and 4.16(b) provide a visualisation of the modified RWP model for a mobility scenario of 10 MN, illustrating intermediate MN way-points and their respective trajectories.

The modified RWP model allows a relatively small mobility transience which in [275] is found to be 300sec. To this end, our simulations adopt a transient period of

400sec and a running period of 1500sec. During the transient period no protocol interactions[23] are initiated and no statistics are collected. To provide a measure-independent view of performance in observed statistics results normalised within the interval [0,1].



(a) PoA topology with MN Way-points

(b) MN Way-points and Trajectories

Figure 4.16: Distributed way-points and trajectories for movement scenario of 10 MN (isolated steady state period of 1500sec) with pause of 10sec running at max speed of 10m/s

Where influence of speed is investigated a fixed pause period of 10 seconds is selected. Where the influence of pause period is investigated a speed of 10m/s is configured. In scenarios where the number of PoA (AR/AP) is kept fixed, simulations encompass network topologies of 45 PoAs. For scenarios where the number of MN is kept fixed, the movement of 10 MNs is investigated.

In the case of speed-influenced scenarios, speeds of up to 90 m/sec are explored. For pause-influenced scenarios, pause periods of up to 50sec are investigated.

4.9.5 Speed model

In the random way-point model employed we extend the generation of average speed value between two way-points by introducing the concept of *micro-trip*. We define as micro-trip the excursion distance between two successive time loci at which the vehicle, cycle or pedestrian has negligible velocity [24] as shown in Figure 4.17(a). The set of micro-trips comprise a single movement path between two randomly generated way-points.

The notion of micro-trips is essential for the purposes of modelling unexpected (i.e.

[23] since simulation need to capture the rate of HARD convergence.

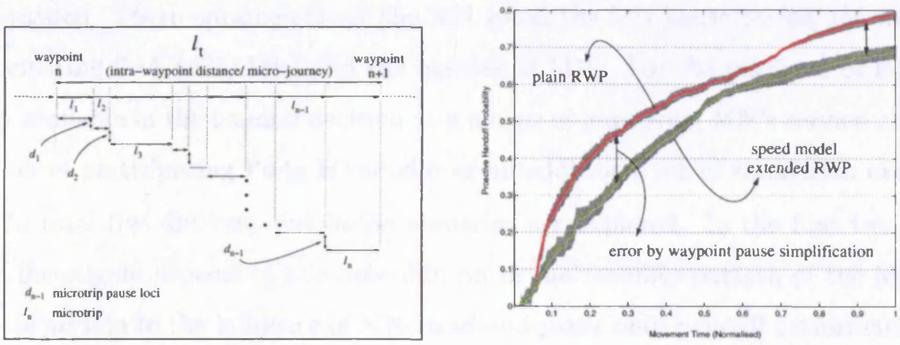
[24] Stationary or near-stationary vehicles have negligible velocity

random) stops, providing a more realistic movement pattern; for instance cars in their movement along two way-points (such as two city locations along their journey towards a final destination) are expected to stop with some rate as a result of crossings or traffic lights.

In the conventional RWP model the node moves between two random way-points without pause. Our extensions introduce a number of k micro-trips (i.e. $k-1$ pauses of n seconds) between two consecutive way-points. The size n_i of each pause period is uniformly distributed between $[0, P_{WP}]$ seconds where P_{WP} is the pause period of the *next* way-point. The number of micro-trips follows a uniform distribution between [1,5].

By extending the RWP model with the revised speed-pause model devised our simulations achieve an increase in accuracy in derived results by about 7-11% as shown in figure 4.17(b) .

The notion of micro-trip is analog to the notion of *sequence* used in [278, 279] for vehicle emissions' measurement in transport research.



(a) RWP micro-trip extension

(b) Relative accuracy gains by micro-trip extension

Figure 4.17: Random way-point speed model extension by introducing micro-trips between way-points. The plot to the right illustrates the relative gain in accuracy by introducing micro-trips between successive way-points for an MN with max speed of 6m/s.

We condition further the average speed estimate $\overline{V}$ by the delay d_{n-1} between consecutive micro-trip units l_{n-1}, l_n. The sum of micro-trips and the intermediate delay components are combined to derive the average speed estimate from equation 4.15

$$\overline{V} = \frac{l_t}{\left[\sum_i \frac{l_i}{v_i} + \frac{\sum_j d_j}{3600}\right]}(km/h) \tag{4.15}$$

Where l_i is the length of the i^{th} micro-trip (km/h), whereas l_t is the total length of the distance crossed until the next *macro-trip* renewal period; d_i is the pause period between consecutive micro-trips and v_i is the MN's mean speed (km/h) at the i^{th} micro-trip of the total distance travelled between two consecutive random way-points.

Equation 4.15 is simplified to:

$$\overline{V} = \frac{l_t}{\left[\sum_i \frac{l_i}{v_i} + \sum_j d_j\right]}(m/sec) \tag{4.16}$$

where both $\overline{V}$ and v_i are now measured in m/sec. For pedestrians and bicycles v_i is assumed to be a fixed value [25] for all micro-trips.

Simulation Parameters and Metrics

In the simulations conducted the influence of four random variable onto handoff delay is investigated. These parameters are the MN *speed*, the MN *pause period*, the *number* of participating PoA (AR/APs) and the *number* of MNs. For the purposes of PoA selectivity available in the handoff decision as a means of improving MN's service utility, the number of participating PoAs is varied over an additional set of simulation executions.

In total five different simulation scenarios are explored. In the first two, simulations investigate aspects of non-determinism in the mobility pattern of the MN; these aspects pertain to the influence of MN *speed* and *pause* onto handoff performance under proactive handoff management. For the third scenario, simulations explore the effect of varying spatial *PoA density* within a geographical region. In the fourth one, simulations explore the effect of varying *MN density* within a geographical region, onto handoff performance achieved under proactive handoff management.

Each simulated PoA (AR/AP) topology encompasses a minimum of 45 PoAs with 45 ARs comprising the greater Routing Neighbourhood and an equal number of APs comprising the Mobility Neighbourhood topology. The number of PoAs is varied between 45 and 1000, while the number of MN is varied between 10 and 1000.

[25] given that it is relatively small and acceleration or deceleration is relatively fast and thus does not affect significantly the average speed estimate

A similar number of PoAs is employed when investigating the scenario of PoA diversity for the purposes of proactively managed handoff selectivity to support increased service utility with respect to MN's service utility criteria or policy. All MNs are communicating with a single CN a simulated VoIP stream through a CBR UDP flow of 13.2kbps (GSM encoding) with 33 bytes as talk-spurt payload. The rationale of not opting for an on/off VBR flow is the fact that the MN during its off-period (silence upstream) is receiving downstream (CN on-period) VoIP packets from the CN. In addition, the simulation scenario assumes no silence suppression during VoIP communication for the purposes of identifying the upper bound of packet loss during an IPv6 handoff and a comparative measure against VAD application in future work.

In terms of simulated IPv6 protocol characteristics for reactive MIPv6, the average Router Advertisement period is set to 40ms, while for Duplicate Address detection (DAD) a constant delay of 1000ms is configured, both according to [107]. The *root-leaf* intra-domain end-to-end delay is set to obey a shifted gamma probability distribution function (P.D.F.) with range between 25-47ms.

For RTT calculation purposes, CN-MN evaluated RTT values are synthetically generated from root-leaf end-two-end delays and confirmed against an exponential distribution[26] with range between 80-160ms according to the discussion of Section 2.3.3. The simulation encompasses also an exponentially distributed measure of *link-local* hangover delay following a heavily-tailed (99th Quantile =350ms) gamma distribution with a mean of 80ms, similar to the one derived from experimental results of Table 3.2, Section 3.7.1.

Table 4.3 provides a complete view of the main simulation parameter configuration employed throughout the performance evaluation of proactive MIPV6 handoff management against its reactive MIPv6 counterpart. Figures in parentheses inform about the fixed values adopted in simulation runs where the influence of different random variables was evaluated.

To attain a reasonable confidence interval each simulation is executed for 20 runs, while statistics are post-processed in Matlab. Where necessary 90, 95 and/or 99% confidence intervals are attained. The spread of the three individual confidence intervals allows to attain higher confidence on the mean (or median[27]) value observed statistically from simulation traces.

---

[26] special case of the gamma P.D.F.

[27] where necessary

Sim. Parameter	Value	Sim. Parameter	Value
Network topology and Coverage and mobility characteristics			
Speed (m/s)	1-90 (10)	Grid size (m^2)	450x450
Pause (sec)	0-50 (10)	Mobility Model	modified-RWP
No of PoA	45-1000 (45)	Transient period	300 sec
No of MN	10-1000 (10)	Steady-state period	1500 sec
VoIP flow characteristics			
packet flow	bi-dir CBR	VoIP sim. encoding	GSM (13.2 kbps)
packetization rate	20ms	packet size	33 bytes
VAD	no		
IPv6 protocol characteristics			
Avg Rt. Adv Interval	40ms	DAD delay	1000ms
e-2-e delay type	shifted Gamma P.D.F	RTT delay	shifted Expon. P.D.F
-//- range	25-47ms	-//- range	80-200ms
hangover delay type	shifted Expon. P.D.F		
-//- range	300-400ms		
MAC layer protocol characteristics			
AP Tx range	10-70m (70)	AP power (mW)	50
Beacon Freq	100ms	Min AP overlap	10m

Table 4.3: Simulation execution parameters at different layers

The simulation assumes infinite MN capacity at each PoA (AR+AP) for both mobility protocols, and unless otherwise stated, an error-free channel with no collisions.

The metrics investigated during statistical analysis are:

- : *handoff delay*: the amount of delay between detachment from the previous PoA until receipt of the first packet at the new PoA

- : *jitter*: the amount of incurred delay variance between consecutive packet arrivals as a result of an IPv6 handoff.

- : *packet loss*: the amount of packets lost as a result of handoff delay. The value is dependent on the packetization rate and is thus codec-specific. In the case of GSM codec the packetization rate is 20ms. Given that smaller (10ms) or larger (30ms) packetization rates can exist for different VoIP codec, the choice of the GSM encoding is representative of average packet loss behaviour for the typical (default) packetization rate in most VoIP systems.

- *reactive handoff probability*: the probability of a reactive handoff once proactive handoff AR discovery (HARD) has been initiated.

- *proactive handoff probability*: the probability of a proactive handoff once the HARD mechanism has been initiated

- *reactive handoff decay*: the measure of decay in the frequency of reactive handoffs as a result of the introduction of the HARD mechanism under proactive handoff management
- *proactive handoff gain/efficiency*: the measure of gain/efficiency in handoff performance from the introduction of proactive handoff management and handoff AR discovery.
- *reactive/proactive probability moving averages*: the rate of decrease in reactive handoff probability and rate of increase in proactive handoff probability as a result of introducing the proactive handoff management.

Randomness of MN movement

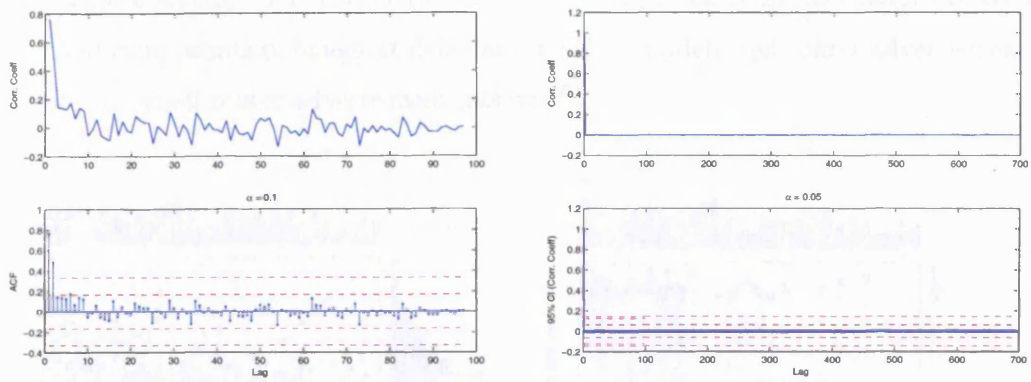
Randomness amongst the mobility patterns of individual MN is essential to ensure that state convergence of protocol performance is not dependent on idealised (but unwanted) event syncronisation (e.g. all MN follow synchronised stops but different trajectories). Randomness in pause occurrence (i.e. not the measure of the pause period) is important to ensure avoidance of synchronised pause (phase) effects which can clock the performance of HARD convergence and thus affect the accuracy of our performance evaluation.

To ensure randomness in the generation of movement and speed patterns for each MN, we look into the sample autocorrelation function (ACF) of inter-movement (assignment of X,Y coordinates), inter-microtrip pause and inter-speed distance (i.e. lag) with the introduction of the microtrip and speed stationarity. Figure 4.18(a) presents the derived correlation coefficient and its respective 90% confidence intreval of the inter-microtrip pause distance. The autocorrelation plots for inter-movment distance is shown in figure 4.18(b); autocorrelation of inter-speed distance exhibits identical behaviour with the one of figure 4.18(b).

It can be seen that the distance between individual pauses shows little to no correlation, validating the simulation expectation that MNs (e.g. encountering traffic lights) will pause randomly. Similar observations apply for inter-movement/speed distance.

4.10 Simulation Results

The first set of simulation results are derived for the simple scenario of 10 MN distributed across the mobility grid roving with a maximum speed of 10m/s, pause of 10sec over a topology of 45 PoA. This scenario provides sufficient and necessary evidence about



(a) inter-microtrip pause ACF (b) inter-movement ACF

Figure 4.18: Very low autocorrelation for inter-microtrip pause and zero autocorrelation for inter-movement distance ensures that mobility sample is truly random

the superiority of proactive handoff management in terms of handoff delay, jitter and packet loss, under nominal speed and pause measures for a small number of MN over its reactive handoff MIPv6 management counterpart. This scenario represents also the nominal case of proactive handoff management performance in terms of handoff delay, which is subsequently compared against scenarios involving a varying measure of speed or pause period as well as varying number of PoA or MN.

All traces collected were subject to statistical filtering for the purposes of outlier removal. Annex E.7 presents in detail the methods employs for statistical identification and removal of extreme values from collected traces.

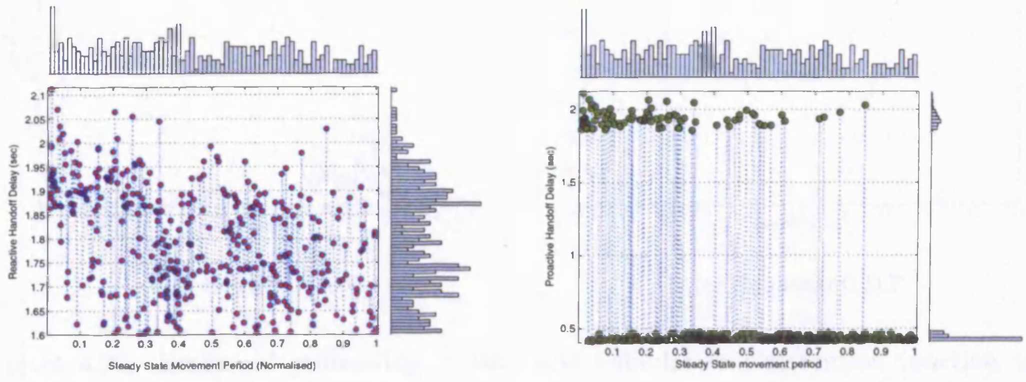
4.10.1 Comparative Handoff performance

Having removed 'outlier' data that can affect the tendency of derived statistics, we now proceed to evaluate statistically the performance of proactive handoff management versus its reactive counterpart MIPV6 standard.

Figures 4.19(a) and 4.19(b) show the comparative handoff delay performance between a standard reactive and a proactive MIPv6 handoff. In the reactive case, handoff delay experiences a high scatter between 1.614 and 1.9 sec with very infrequent ($< 1\%$) handoff delays over 2sec. The reactive handoff case exhibits a bi-modal pattern with delay peaks of 1.69sec and 1.86sec not exceeding 8.2% and 10.8% respectively. This is evident from the empirical probability distribution presented on the vertical[28] side

[28]the horizontal probability distribution shows the average handoff frequency for the simulation scenario.. It demonstrates a uniform probability and thus validates the randomness of mobility pattern among MNs.

of the Figure 4.19(a). The two modes are attributed primarily in fluctuations of RTT delay and components of hangover delay as a result of undetected router advertisments (despite the small router advertisment interval[29]).



(a) Reactive Handoff Delay (MIPv6)   (b) Proactive Handoff Delay (PMIPv6)

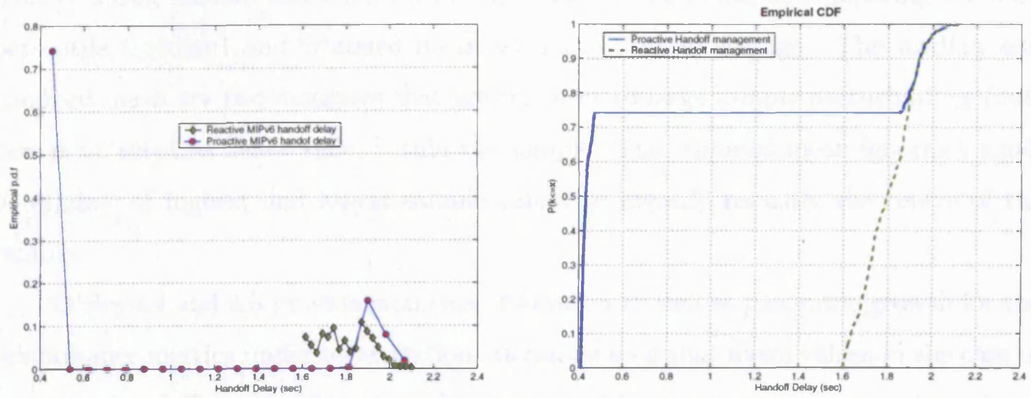
Figure 4.19: Handoff delay experience during a standard reactive and a proactive IPv6 handoff

For the proactive handoff case the proactive handoff component occurs with a probability of 0.742 incurring (see vertical distribution plot of Figure 4.19(b)) a delay of less than 455ms with about 380-420ms attributed to delay incurred by the link-layer handoff. The reactive handoff component while decaying during HARD convergence exhibits a probability of 0.157 with a delay of 1.89sec. Less than 2% of the reactive handoffs note a delay of over 2sec and about 8% experience a delay of 1.98sec in this IP handoff component.

Figures 4.20(a) and 4.20(b) present the empirical probability density and cumulative distribution functions of the observed measure of handoff delay performance between reactive and proactive MIPv6. From the probability density it may be seen that reactive MIPv6 concentrates delay around high figures of 1.6-1.9 sec but is distributed over smaller (< 10%) probability densities. On the contrary, proactive MIPv6 *polarises* handoff delay performance between the proactive and the reactive handoff delay component, with a decaying peak for reactive handoff delay transformed into a growing peak for proactive handoff delay.

The decaying reactive part appears by about 5% more probable that one in a normal reactive handoff for the same reactive handoff delay values. The difference is justified by the use of different *bin ranges* in capturing the probability density of small

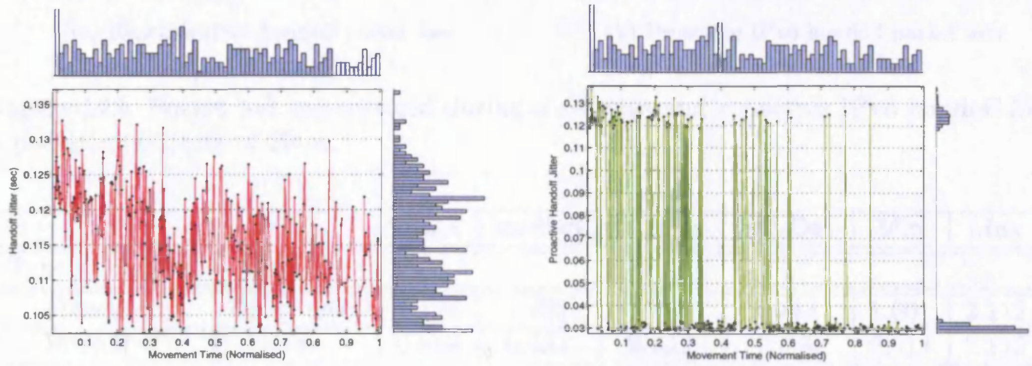
[29]See conclusions pertaining to hangover delay in Chapter 3



(a) Empirical p.d.f (b) Empirical C.D.F

Figure 4.20: Empirical probability density and cumulative distribution function of proactive versus reactive MIPv6 handoff delay

reactive handoff ranges in the reactive MIPv6 handoff performance case. That is to say, the density of reactive handoffs is calculated through a smaller handoff delay range to capture with higher granularity the effect of handoff delay fluctuations between small ranges. On the contrary, the probability density of proactive handoffs is collectively computed over larger bin ranges since handoff delay values vary much less than in the reactive handoff case.



(a) Reactive IPv6 handoff jitter (b) Proactive IPv6 handoff jitter

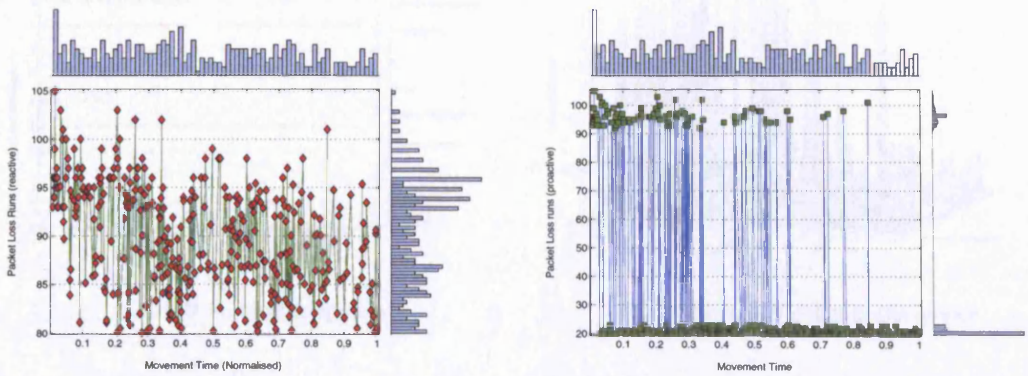
Figure 4.21: Jitter experienced during a reactive and proactive IPv6 handoff

To identify any positive or negative skewness[30] of the data as a result of extreme handoff delay values, we derive the basic measures of location (aka central tendency,

[30]skewness in statistical distribution curves or percentiles or difference between mean and median values, indicates deviation from normality, and distribution tails, i.e. values that are fairly infrequent ($\leq 1\%$).

namely mean, median and trimmed mean). This can be found by comparing the 50th percentile (median) and trimmed mean with the sample average. The median and trimmed mean are two measures that are resistant to large sample perturbations (outliers manifested as heavy tails) within the sample. The trimmed mean ignores a small percentage of highest and lowest sample values to identify robustly the centre of the sample.

Tables 4.4 and 4.5 provide moments of location as well as percentile growth for the performance metrics under investigation. It can be seen that mean values in the case of proactive handoff are significantly right-skewed and hence are not representative of avg performance. For this reason we consider the *median* value as the statistic exhibiting a meaningful measure of location.



(a) Reactive IPv6 handoff packet loss

(b) Proactive IPv6 handoff packet loss

Figure 4.22: Packet loss experienced during a reactive and proactive IPv6 handoff for a packetization rate of 20ms

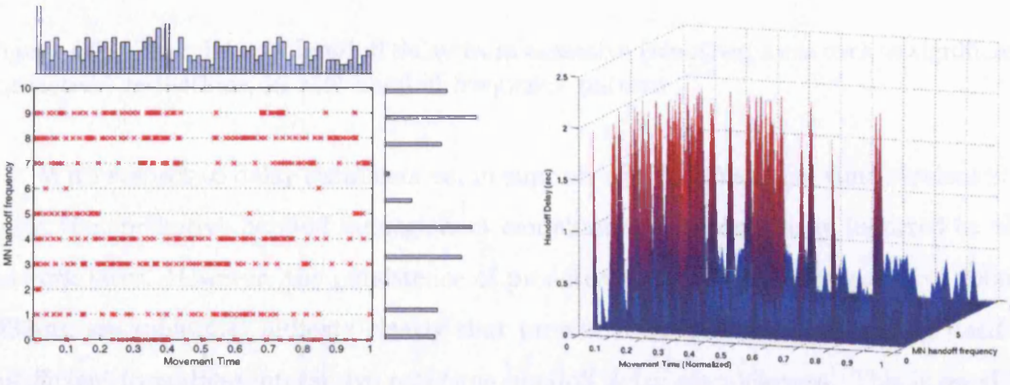
Type of handoff	Metric	mean	median	trimean	Std.Dev	Min	Max
Reactive	Delay (sec)	1.798	1.794	1.797	0.114	1.601	2.112
Proactive	-//-	0.819	0.433	0.693	0.660	0.413	2.112
Reactive	Jitter (sec)	0.115	0.114	0.115	0.007	0.102	0.136
Proactive	-//-	0.054	0.030	0.046	0.041	0.028	0.136
Reactive	Pkt loss ()	89	89	89	5.57	80	105
Proactive	-//-	40	21	33	33	20	104

Table 4.4: First statistical moments indicating central tendency in the set of simulations for nominal speed and pause of a sparse set of (10) MN distributed within a sparse topology of 45 PoAs

Figures 4.21 and 4.22 present the respective measure of jitter and packet loss as a result of the aforementioned handoff delay measure. It can be seen that proactive

		Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
Type of handoff	Metric						
Reactive	Delay (sec)	1.705	1.892	1.94	1.979	2.052	2.112
Proactive	-//-	0.42	1.765	1.84	1.89	2.011	2.092
Reactive	Jitter (sec)	0.109	0.121	0.125	0.127	0.137	0.139
Proactive	-//-	0.029	0.119	0.121	0.125	0.132	0.135
Reactive	Pkt loss ()	85	94	96	98	102	105
Proactive	-//-	21	93	96	97	101	104

Table 4.5: Percentiles indicating the cumulative growth of observed metrics in the set of simulations for nominal speed and pause of a sparse set of (10) MN distributed within a sparse topology of 45 PoAs

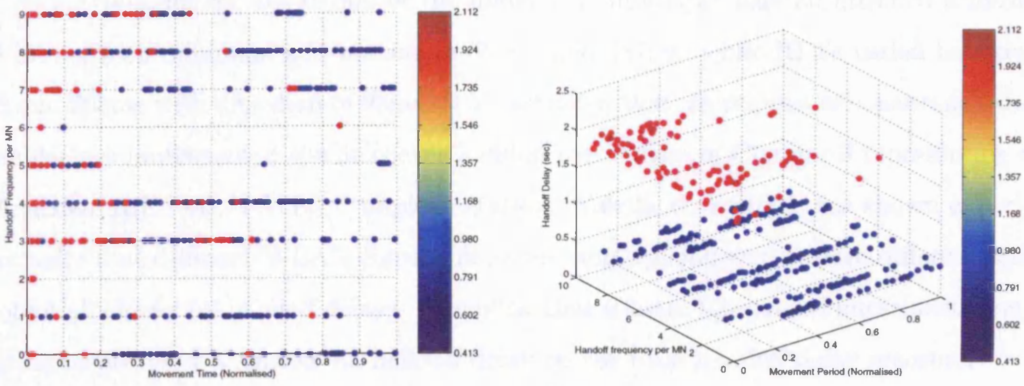


(a) MN handoff frequency (b) MN handoff delay performance

Figure 4.23: Handoff frequency of the simulated 10-MN set and their respective handoff performance in terms of Proactive (blue) versus reactive (red) MIPv6 handoff delay

handoff management reduces both jitter and associated packet loss to about 20% of the respective measure incurred by reactive MIPv6 handoff management. This is an improvement of about 80% in handoff delay, jitter and packet loss, once handoff AR discovery state has converged. The probability density function of figure 4.20(a) reveals further, that such magnitude of improvement degrades by *only* about 5.79% with respect to the total movement period, while HARD state has not reached its convergence threshold.

Given that not all MNs are necessarily engaging in seamless communications simultaneously while in transit, it becomes easy to see that, the nearly 6% degradation on handoff performance as a consequence of HARD state convergence, can quickly be eliminated by non-communicating MNs simply in transit. Note that the aforementioned measures apply for the nominal case of a small number of MN (10), at low MN speed (10m/s) and relatively small pause period (10sec).



(a) 2-d view (b) 3-d view

Figure 4.24: Transition of handoff delay from excessive (reactive) measures to significant (proactive) reductions, in MN handoff frequency pattern

With respect to delay seamlessness, in support of interactive real-time services it is found that proactive handoff management can eliminate handoff delay incurred by the network layer. However, the persistence of proactive handoff delays in excess of 200ms (433ms, see table 4.4) indicate clearly that proactive handoff management by itself is insufficient to address interactive real-time handoff delay seamlessness. This is owed to two important delay influences: (i) external latency factors such as increasing round trip times (RTTs) and (ii) link-layer handoff delay. These are the two sole components that persist in affecting adversely seamless handoff delay performance.

However, different wireless technologies implement a link-layer handoff through different approaches[31] and thus, incur significantly disparate measures of L2-handoff delays [280]. In addition, fundamental requirement for successful Internet system design, is the sustainment of strict layer independence with no assumptions about the type of L2 technology.

In line with the above, the most appropriate mechanism of attacking external latency and L2-handoff delay intrinsics appears to be proactive flow management and buffering at the network layer during the period of the handoff. We postpone any analysis on proactive flow management until Chapter 5.

It is, however, important to note that in the vast majority of the observed proactive handoffs, handoff delay performance is found to be *directly dependent on the period of completion of the underlying L2-handoff*, rather the measure of RTT. In the majority

[31] For instance, in cellular systems native IPv6 signalling operates over the legacy GPRS signalling before a single control handshake takes place, inflating the delay total.

of proactive handoffs, the period of the underlying link-layer handoff attained a mean of 375ms with minimum and maximum of 365 and 487ms, while RTTs varied between 58 and 205ms with a median of 93ms. It is reminded that for reasons of consistency the simulations implemented the link-layer handoff mechanism of Chapter 3 representing a particular IEEE802.11 WLAN implementation. Yokota et al [105], has shown experimentally that different WLAN chipset/firmware implementations achieve different and potentially lower L2-handoff delays. It implies, that a faster L2-handoff implementation, incurs respectively a smaller L2-handoff delay on the total handoff delay measure.

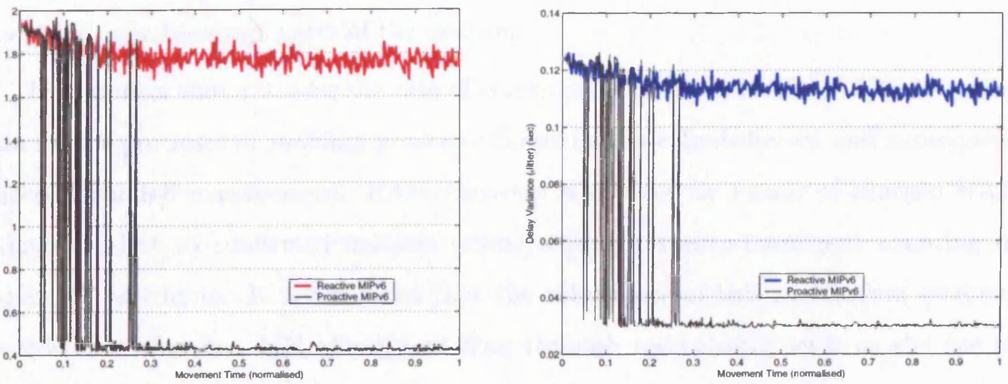
4.10.2 Dependence on HARD state convergence

To attain a better understanding of handoff delay performance as a function of the total number of handoffs spread across the set of MNs we expand the movement trajectory of MNs into their respective handoff frequencies through a random simulation run. Figure 4.23(a) shows the frequency pattern of handoffs from individual MNs during a single simulation run, while the respective histogram (axis y) indicates the handoff distribution; The handoff sequences for each MN indicate that wireless hosts perform with different frequency their PoA transition, but do not indicate at which point each MN begins to differentiate between proactive and reactive handoffs in support of handoff delay seamlessness.

Figure 4.23(b) expands this view with a (colour) mapping of handoff delay figures, visualising the time period that HARD reaches convergence to support robustly proactive handoff management, with significant reductions in observed total handoff delay. The visualisation is enhanced in figure 4.24. The 2-d plot view presents visually the rate of transition from reactive (red-oriented colormap) to proactive (blue) handoffs. For most MNs the first few handoffs are experienced with reactive-type delays.

However, a number of these MNs begin to experience proactive-type handoff delays almost right from the beginning of their movement pattern. This is better illustrated in the 3-d view where handoff delay values are plotted against axis z to differentiate according to their magnitudes, in addition to MN identity and simulation time. It can be seen that for that measure of speed and pause time, at about 0.4-0.5 of the simulation period (around 600sec) reactive handoffs occur with very low probability indicating some reasonable measure of HARD state convergence.

Figures 4.25(a)-4.26 present the comparative performance in terms of handoff delay, jitter and the associated packet loss between proactive handoff management and its reactive MIPv6 counterpart. It can be seen that proactive handoff management maintains a distinct performance advantage in all three metrics, while being temporally dependent on the rate of convergence of HARD state. To this end, the following section elaborates on the measure of HARD state convergence and parameters of influence.



(a) Handoff Delay

(b) Jitter

Figure 4.25: Handoff delay and associated jitter under Reactive (red/blue) and Proactive (black) IPv6 handoff management. The high-valued colour curve indicates measures of delay and jitter under standard reactive MIPv6. The measure of reactive handoff delay and jitter is nearly constant while the respective measure under Proactive MIPv6 reduces dramatically as HARD state reaches convergence.

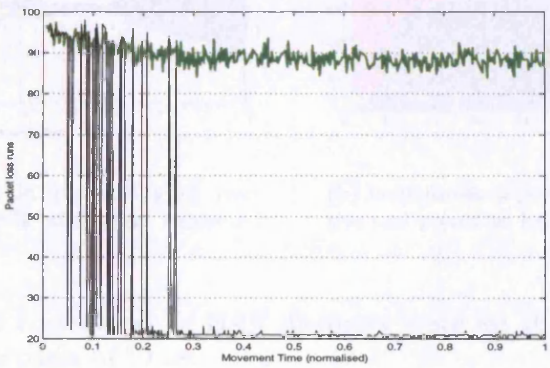


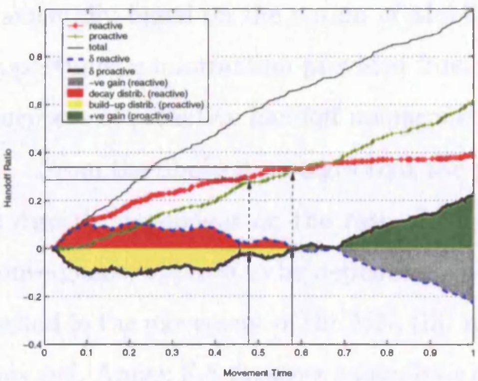
Figure 4.26: Associated packet loss runs under Reactive (green) and Proactive (black) IPv6 handoff management.

4.10.3 HARD state Convergence

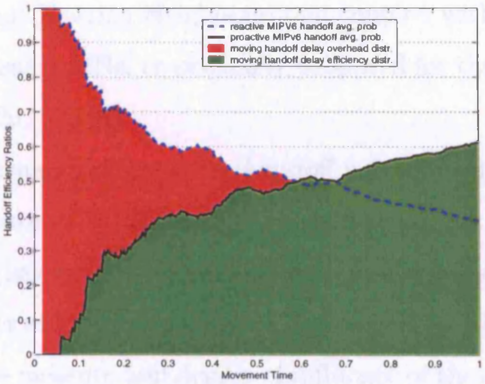
The previous section has elaborated on the measure of total handoff delay. It has illustrated that the rate of transition from reactive to proactive handoff delay measure is effectively dependent on the rate of Handoff AR discovery (HARD) state convergence.

With respect to the type of handoff AR update algorithm employed, both results and analysis focus on the charged HAR update, since the boosted charged update did not influence significantly the completion of the handoff TMM matrix at each AR neighbour. Its fill ratio appears to become important only in the case of proactive flow forwarding management elaborated in Chapter 5. Furthermore, plain hinted HAR discovery updates exhibit approximately half the performance in terms of the metrics elaborated in subsequent parts of this section.

This section elaborates on the rate of convergence of handoff AR (HAR) discovery state for the purposes of enabling proactive context state establishment and subsequent proactive handoff management. HAR discovery is effected by means of charged HAR updates subject to confirmed updates (three) through swarm-intelligent counting of by-passing MN hints. It is reminded that the robustness of this mechanism assumes the existence of secure MN identity/naming through mechanisms such as the use of cryptographically generated IPv6 addresses [281].



(a) Rate of growth in probability of reactive/proactive handoffs and their respective overhead/gain



(b) Instantaneous probability measure of reactive and proactive handoff

Figure 4.27: Rate of convergence of HAR discovery state for 10 MN moving at max speed of 10 m/s with pause of 10 sec

Figure 4.27(a) shows the rate of growth in probability of HARD state convergence manifested as transition rate probability from reactive to proactive handoffs. At approximately 0.5 of the movement time the buildup of HARD state causes the rate of growth of proactive handoffs to overtake the rate of reactive handoffs which is steadily decreasing, settling the reactive handoff rate probability to about 0.3. After time 0.6 (900sec) the rate of growth reactive handoffs becomes nearly-constant, translating to a $< 1\%$ probability of reactive handoffs occuring in the system. This is better illustrated by the reactive handoff decay distribution of figure 4.27(a) and the average decay probability of figure 4.27(b).

Figure 4.27(a) encompasses further the measure of negative gain incurred by reactive handoffs as a result of the rate of HARD state build-up. The is complemented by the symmetric positive performance gain incurred by an increasing number of proactive handoffs as the HARD state cache in neighbouring PoAs reaches convergence. It can be seen that for the particular set of pause and speed parameters as well as the size of PoA topology and the sparseness of moving MNs HARD state among all 45 PoA of the topology requires about 0.47 of the simulation time, despite the steady increase of proactive handoffs shown from the proactive handoff ratio curve.

The validity of these results is further confirmed by preliminary results of an independent study by Chalmers et al [147, 282] on the performance benefits of a dynamic candidate AR discovery (`dyCARD`) mechanism. `dyCARD` is essentially a standardisation implementation effort of the HAR discovery mechanism [47]; its core functions are fundamentally based on the notion of Mobility and Routing Neighbourhood coupled with opportunistic information provided from bypassing MNs, as originally proposed for the purposes of proactive handoff management [129, 50, 49].

From the above it emerges that the performance of proactive handoff management is directly dependent on the rate of convergence of HARD state. Such type of state convergence, appears to be dependent on four key aspects: (i) MN speed, (ii) the pause period in the movement of the MN, (iii) number of MN (iv) density of PoA topology. To this end, Annex E.8 presents an analysis on the measure and degree of influence of these factors on HARD state convergence for the purposes of proactive handoff management.

4.10.4 Discussion and Results Summary

With respect to the observed measure of proactive handoff delay the simulation study has shown that such delay performance is directly related with the availability of proactively established IP Roaming state. The latter is essentially dependent on the convergence of handoff AR discovery (HARD) state for subsequent state establishment.

Convergence of HARD state is in turn influenced by the measure of non-determinism in the mobility pattern of the MN. Section 4.10.3 has shown that the rate of convergence of HARD state effectively introduces a frequency of oscillation between reactive and proactive handoff delay performance as HARD state builds up at HAR neighbours. Past this convergence period proactive handoff delay improves significantly guarantees towards hard delay bounds for the purposes of interactive real-time communications. However, results show that it becomes imperative that the HARD state convergence period remains short, in comparison to (the simulated) MN movement period if proactive handoff management is to maintain a characteristic advantage compared to reactive MIPv6 handoff management. Furthermore, initial results showed that the measure of HARD state convergence period is directly influence by the stochastic nature of MN's movement pattern.

To this end, we identified key parameter describing aspects of non-determinism in the mobility pattern of the MN; these were the measure of MN speed, pause period between way-point stops, the number of MN within a nominal PoA topology and the measure of PoA density describing the physical PoA (AR+AP) network topology. To this end we devised a detailed set of simulation scenarios whereby the performance of each of these parameters brought into scrutiny and analysed, while background parameters retained nominal values. This allowed features of each parameter under investigation to reveal their behavioural pattern.

With respect to the measure of influence of MN speed in the convergence period of HARD state, results of Section E.8.1 show significant differentiation in proactive handoff delay performance. In particular, results have shown that for very low maximum MN speeds (e.g. pedestrian) of 1m/s HARD state convergence periods are prohibitively slow giving rise to a consistently higher reactive handoff probability and sustained frequency of oscillations between reactive and proactive handoffs throughout the simulation period. On the contrary, the probability of proactive handoff improves dramatically for maximum MN speed measures above 6m/s. That is to say, HARD state convergence improves when the set of participating MNs adopt a considerably heterogeneous MN speed mix. Simulation results show further that the probability of a proactive handoff does not increase linearly with the measure of speed. For speeds $> 30m/s$ (highway), the probability of a proactive handoff improves only marginally (by 6%), while a maximum speed of 12m/s appears to bring the best trade-off between MN attainable speed and HARD state convergence. Similar results are reported by Chalmers et al [147], by showing a similar convergence rate of dyCARD state over smaller velocity measures.

With respect to the measure of influence of pause period in the movement pattern of the MN results have shown that for pause periods between 5-50 sec the probability of a proactive handoff through HARD state convergence is affected only by 13-15%. For a 10-fold increase (6-60m/s) on the measure of maximum MN speed the probability of a proactive handoff improves by 22%. The former implies that pause periods between 5-50sec do *not* affect significantly the rate of HARD state convergence, while any influence can be quickly compensated by a higher maximum measure of MN speed amongst nodes. The latter is readily attainable under real-world operational conditions since, MN at different parts of the PoA topology experience typically highly heterogeneous measures of maximum MN speed.

With respect to the measure of PoA density within the topology, results show that as PoA density increases from 45 to 200 the probability of a proactive handoff decreases for a sparse set of MNs (10) by about 9%. Interestingly enough however, for PoA topologies above 200 (200-1000) the probability of a proactive handoff remains nearly constant with a near-identical rate of increase in HARD state convergence. This is justified by a significant increase in the instantaneous handoff rate, since the PoA density/m^2 increases significantly.

The significant increase of instantaneous handoff rate however reveals a significant detail in the performance of the simulation model. The observed performance for the particular scenario assumes stateful generation of IP Roaming state and thus reduces the minimum measure of MN cell residence period, such that a higher (sub-second) handoff rate can be attainable. If the minimum cell residence period is enforced for the MN, then for instantaneous handoff rates > 1 handoff/sec the MN cannot attain proactively IP Roaming state in a proxy-stateless fashion; this is so because, proxy-stateless address auto-configuration would require a minimum of 1sec for DAD resolution purposes at the HAR neighbour. As a result of unavailability of proactively established IP Roaming state, the MN would need to resort to a reactive handoff. However, Chapter 3 has shown that reactive MIPv6 is unable to perform for handoff rates greater than 0.5 handoffs/sec. The above lead us to conclude that if the instantaneous handoff rate exceeds the minimum cell residence period, neither proactive nor reactive MIPv6 handoff management can handle the mobility pattern of the MN. Under such operational scenarios the MN cannot receive realistically any form of IPv6 mobility management support. The resulting requirement emerging from these results is that the instantaneous handoff rate of the MN *must* remain less than the minimum measure of cell residence period if (i) IPv6 mobility management services are to be supported (ii) proactive IPv6 handoff management is to sustain operational benefit.

With respect to the measure of MN population size results show that for an increasing number of MNs the probability of a reactive handoff decreases dramatically (down to 7%) early during MN movement. The measure of decrease is significantly higher than that attained by a high measure of MN speed indicating that large MNs with moderate maximum speeds allow HARD state to converge in less than 0.01 of the simulation time (i.e. in less than 15sec). As the number of MN increases between 25-200 the probability of a proactive handoff increases monotonically with a proportional rate of increase.

However, results show further a counter-intuitive observation, whereby as the number of MN increases beyond 200 (between 300-1000) the rate of increase in proactive handoff and the rate of decrease in reactive handoff probability are *reversed* during the start of the simulation. This is owed to the large instantaneous rate of reactive handoffs per unit time, experienced by PoAs at the start of the simulation. During this period (which shortens as the number of MN increases) PoAs experience a handoff rate that is proportional to the MN population density/m^2 (see section E.8.4). The effect diminishes when the density of MN per m^2 is decreasing.

From all four parameters of influence MN population size and MN speed appear to incur the highest (positive) rate of convergence to HARD state. MN pause periods and PoA density have a less positive influence, although even for significant increases in their measure, they do not impact the measure of proactive handoff probability and HARD state convergence significantly.

4.11 Capability notification and PoA handoff diversity

To assess the performance benefit of PoA handoff selectivity during execution of MN's IPv6 handoff decision, the simulation model is extended to encompass some measure of heterogeneity in each PoA. Furthermore, to enable a higher handoff rate for MNs operating in dense PoA topologies, the proactive handoff management model is aligned to support explicitly *stateful* address auto-configuration, in line with the findings of section E.8.3.

Heterogeneity amongst visited PoAs is introduced by implementing an abstract capability parameter that differentiates between PoAs through an abstract performance metric. Such metric can be adapted to take numerous IP connectivity-specific forms when applied in a commercial setting, such as, bandwidth, delay, transmission range, or seamless handoff capability. Alternatively, the metric may describe more service-specific capabilities, such as per-minute tariff pricing, emergency information availability, free service features, or other. By adopting an abstract performance metric, the simulation model ensures that results are generalisable across the entire capability spectrum.

The particular simulation model (and its results) focus on *persistent* or *non-volatile* capability metrics, as opposed to *volatile* capabilities. Persistent or non-volatile capability metric is defined to be a quantifiable service characteristic, the measure of which persists throughout the cell residence period of the MN, when associated with the respective PoA. On the contrary, a volatile capability metric is defined to be a quantifiable time-variant service feature, that is, a measure of service capability that changes with time.

With regards to a persistent capability, when communicated proactively to the MN, in advance of the IPv6 handoff transition, persistence in its measure ensures that such capability will be available at that measure when the MN associates with the particular PoA. Examples of non-volatile capability metrics are service features that can be declared with binary (on/off) availability, or with a measure of availability that remains fixed in its provisioning for all (horizontal) or each MN (vertical) individually; these may be data rates, emergency notification services, etc. Exploring the depth of service capability differentiation is, however, beyond the scope of this study and is thus left as a future direction of investigation.

For the purposes of simulation, each PoA is allocated a random measure of the abstract (persistent) service capability with a value ranging in the interval [0,1]. The capability is assumed to refer to a particular service context; hence, capability state pertaining to the particular context is proactively exchanged between individual PoA through direct RNV-updates. These are subsequently pushed to the MN by means of CtS-Response message. During simulations the number of PoA within the movement grid is varied between 45 and 1000 PoA, simulating the scenario of increased number of handoff candidates during MN's next IPv6 handoff. A sparse set of 10 MNs was configured with a maximum speed of 10m/s and a pause of 10sec through the modified random way-point model.

Proactive notification of such context-specific capability metric may be *solicited* or *speculative*. In the case of a solicited proactive capability notification the MN is itself requests such provisioning. In the case of a speculative proactive capability notification, the PoA *pushes* such context-specific information to the MN. The simulation model is configured towards speculative proactive capability notification. Exchange of such proactive mobility management signals during the simulation is configured to be error-free.

4.11.1 Simulation Results

Figure 4.28 shows the measure of attained service utility for the MN by exercising PoA selection during a MIPv6 handoff on the basis of SNR strength information at the link layer. The measure of service utility fluctuates marginally around 0.5 for reactive MIPV6 handoffs that do not exploit the availability of multiple PoA handoff candidates at MN's next IPv6 handoff.

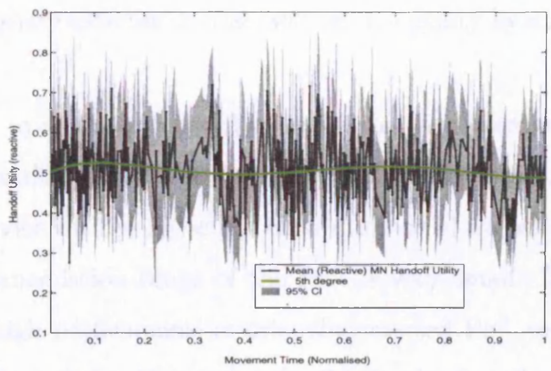
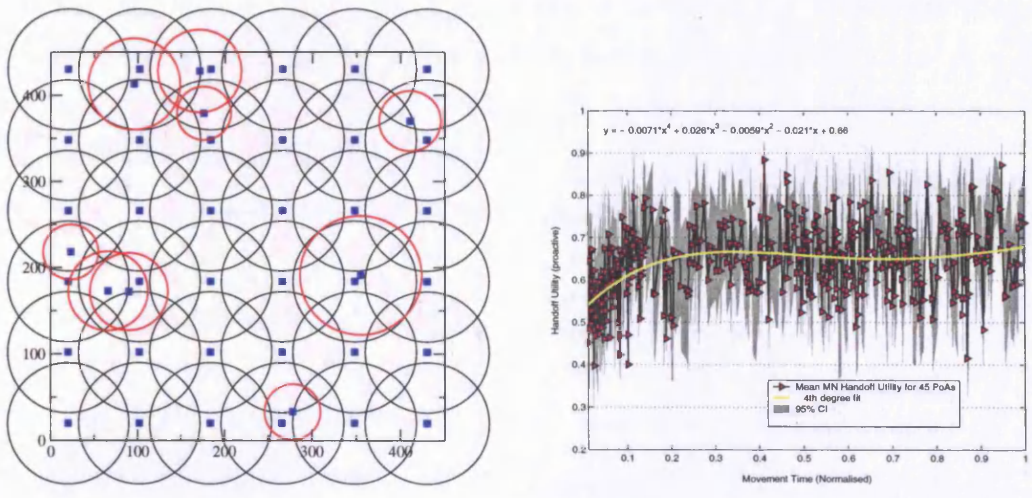


Figure 4.28: Service utility experienced by MNs performing a standard MIPv6 handoff based on traditional SNR strength information.

Figure 4.29-4.32 present the gradual gain in MN's experienced service utility as a result of exploiting PoA diversity by applying PoA handoff selectivity. Such form of handoff selectivity is based on the availability of the highest service capability metric value available within the set of immediate PoA handoff candidates. The criterion for selection in the case of all simulations is the maximum available service metric value supported by one of the PoA handoff candidates.



(a) PoA topology of 45 ARs

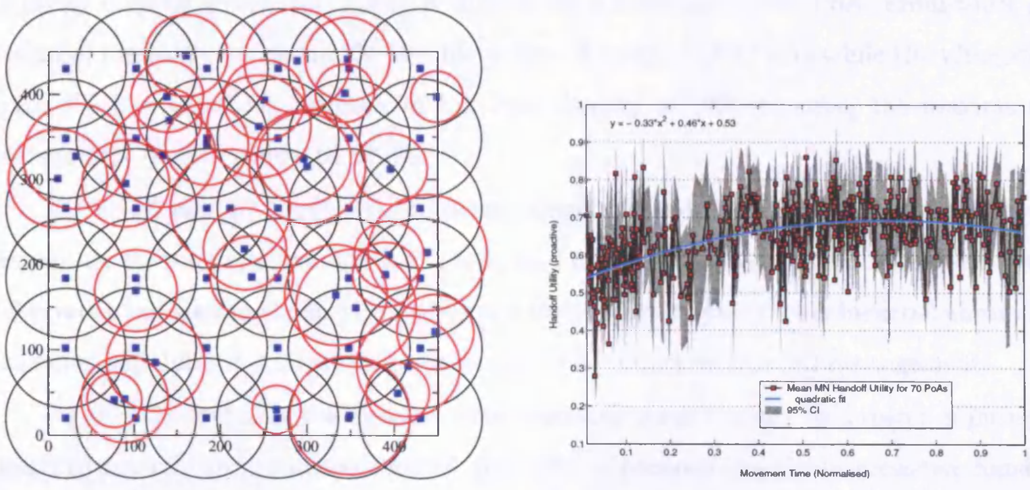
(b) Proactive Handoff Utility @ 45 ARs

Figure 4.29: Average proactive Handoff Utility experienced by 10 MN over a topology of 40 ARs

Figure 4.29 plots the observed measure of service utility contrasted with the respective topology density of 45 PoAs. Out of these 45 PoA, only 9 provide some additional measure of PoA diversity in MN handoff selection randomly distributed over the entire PoA topology. It may be seen from the respective service utility plot that the MN experiences an immediate increase in observed service utility by about 14% as shown in table 4.8.

It is important to note that the fluctuation in the observed measure of service utility arises as a result of *false positive* effect. The MN appears to select the PoA with the highest service utility, although on a number of occasions the cell residence period within the transmission range of the PoA is very small. Reason for this is the fact that despite a high performance metric, the selected PoA serves the MN only at the periphery of its trajectory; that is to say, due to the fact that the selection is not based on SNR information, the selected PoA does not lie in the angle of direction of MN's trajectory.

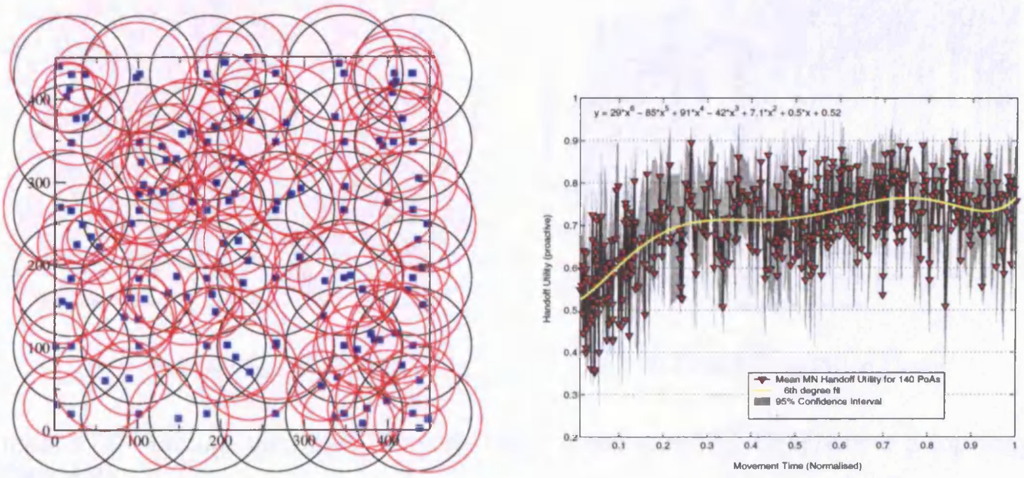
As a result, the association with the particular PoA is short-lived as the MN handoffs very soon to the PoA with next higher service utility metric, which this time happens to have also the highest SNR for the MN. The latter implies, essentially, that strong SNR when combine with high service utility ensures that the selected PoA not only provides the highest performance in terms of the selected metric, but also has a significantly higher probability of coinciding with MN core trajectory.



(a) PoA topology of 70 ARs (b) Proactive Handoff Utility @ 70 ARs

Figure 4.30: Average proactive Handoff Utility experienced by 10 MN over a topology of 70 ARs

Figure 4.30 presents the case for a density of 70 PoAs, whereby PoA diversity of immediate PoA handoff candidates almost doubles (one extra PoA for each of the first 36 PoA providing continuous coverage in the movement grid). Interestingly enough, the average service utility experienced by all MN's increases only by a mere 1%



(a) PoA topology of 140 ARs

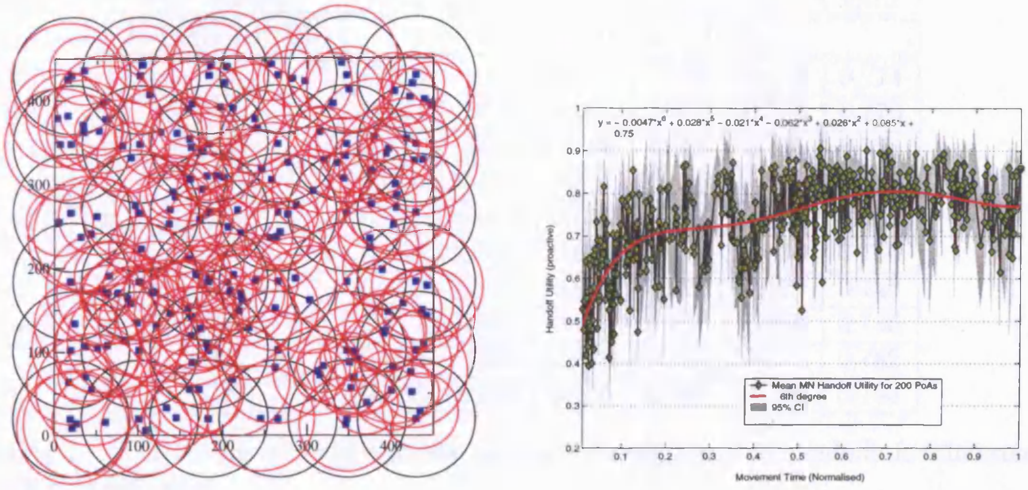
(b) Proactive Handoff Utility @ 140 ARs

Figure 4.31: Average proactive Handoff Utility experienced by 10 MN over a topology of 140 ARs

As the number of PoA doubles from 70 to 140 (see figure 4.31 the average service utility increase by about 5%. From there an increase to 200 PoA increases the measure of utility only by about 3% raising it to 76% for a topology of 300 PoA. From there a marginal increase of an extra 2% is achieve for a density of 500 PoAs while the ultimate extra 2% is reached at the peak of the PoA density of 1000 bringing the maximum average MN service utility to 80.6%.

Tables 4.6 and 4.7 provide the measure of central tendency as well as service utility percentiles for pure reactive handoffs performed under standard MIPv6. Tables 4.8 and 4.9 present the measure of service utility and the respective percentiles incurred through proactive capability notification implemented over proactive handoff management.

Figure 4.33 contrasts the measure of experienced service utility as a result of purely standard reactive and proactive MIPv6 handoffs. It becomes clear that proactive handoff management offers an increase in MN service utility of about 10-30% depending on the PoA density. The maximum service utility value appears to be achieved for a density of 500 PoA with the highest rate if increase at 300 PoA. Between 300 and 500 PoA service utility at the MN increases only by about 2%. From that it may be concluded that the optimum increase at service utility is achieved at 300 PoA.



(a) PoA topology of 200 ARs

(b) Residual of Fitted Curve

Figure 4.32: Average proactive Handoff Utility experienced by 10 MN over a topology of 200 ARs

PoA topology	mean	median	trimean	Std.Dev	Min	Max
# (AR+AP/BS)	Reactive Handoff Utility					
45	0.53	0.527	0.527	0.087	0.302	0.774
70	0.49	0.484	0.486	0.085	0.278	0.776
140	0.487	0.484	0.487	0.085	0.266	0.755
200	0.507	0.508	0.506	0.078	0.276	0.718
300	0.520	0.521	0.520	0.077	0.303	0.733
400	0.510	0.513	0.510	0.075	0.305	0.766
500	0.533	0.531	0.532	0.077	0.305	0.746
600	0.529	0.529	0.530	0.079	0.323	0.737
700	0.519	0.518	0.518	0.082	0.277	0.763
1000	0.534	0.531	0.534	0.080	0.283	0.764

Table 4.6: Moments of central tendency (location) of service utility incurred during reactive handoffs for different PoA topology sizes

Figures 4.34-4.35 plot the respective empirical probability density function of MN service utility experienced under proactive handoff management for PoA densities between 45 and 200 PoA. It may be observed that for a sparse number of PoA in the topology, the probability density is normally distributed around a service utility measure of 0.65. The shape of the distribution becomes steadily right-skewed around 0.8-0.85 for the topology of 200 PoAs.

It is worth noting that again as a result of the selection of bin size and range for exploring probability densities at higher granularity, probabilities peaks register lower values with a higher distribution across the range of service utility configures in the

PoA topology	Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
# (AR+AP/BS)	Reactive Handoff Utility					
45	0.465	0.590	0.650	0.678	0.731	0.774
70	0.433	0.547	0.605	0.636	0.704	0.776
140	0.427	0.547	0.595	0.636	0.697	0.755
200	0.455	0.558	0.610	0.642	0.694	0.718
300	0.466	0.574	0.618	0.655	0.692	0.733
400	0.459	0.556	0.613	0.630	0.680	0.766
500	0.477	0.587	0.632	0.670	0.723	0.746
600	0.474	0.588	0.637	0.658	0.694	0.736
700	0.458	0.575	0.626	0.660	0.712	0.762
1000	0.476	0.589	0.636	0.665	0.720	0.764

Table 4.7: Percentiles of service utility incurred during reactive handoffs for different PoA topology sizes

PoA topology	mean	median	trimean	Std.Dev	Min	Max
# (AR+AP/BS)	Proactive Handoff Utility					
45	0.641	0.649	0.641	0.087	0.397	0.884
70	0.650	0.656	0.655	0.093	0.278	0.860
140	0.699	0.714	0.708	0.103	0.352	0.900
200	0.730	0.747	0.741	0.105	0.394	0.905
300	0.764	0.785	0.774	0.105	0.396	0.95
400	0.767	0.8	0.781	0.121	0.413	0.951
500	0.78	0.803	0.794	0.112	0.411	0.951
600	0.786	0.808	0.799	0.117	0.357	0.968
700	0.787	0.808	0.801	0.12	0.393	0.968
1000	0.806	0.835	0.82	0.125	0.39	0.975

Table 4.8: Moments of central tendency (location) of service utility incurred during reactive handoffs for different PoA topology sizes

selected bins. Annex E provides the remaining set of probability functions of simulated PoA densities larger than 200.

PoA topology	Q25	Q75	Q90	Q95	Q99	Q99.9
# (AR+AP/BS)	Proactive Handoff Utility					
45	0.58	0.703	0.753	0.782	0.844	0.884
70	0.59	0.717	0.760	0.792	0.819	0.86
140	0.637	0.777	0.816	0.842	0.879	0.9
200	0.67	0.813	0.851	0.864	0.893	0.905
300	0.695	0.846	0.882	0.903	0.93	0.95
400	0.689	0.864	0.899	0.917	0.939	0.951
500	0.715	0.862	0.903	0.926	0.941	0.951
600	0.711	0.879	0.925	0.939	0.958	0.968
700	0.721	0.880	0.919	0.944	0.961	0.968
1000	0.733	0.905	0.943	0.96	0.972	0.975

Table 4.9: Percentiles of service utility incurred during proactive handoffs for different PoA topology sizes

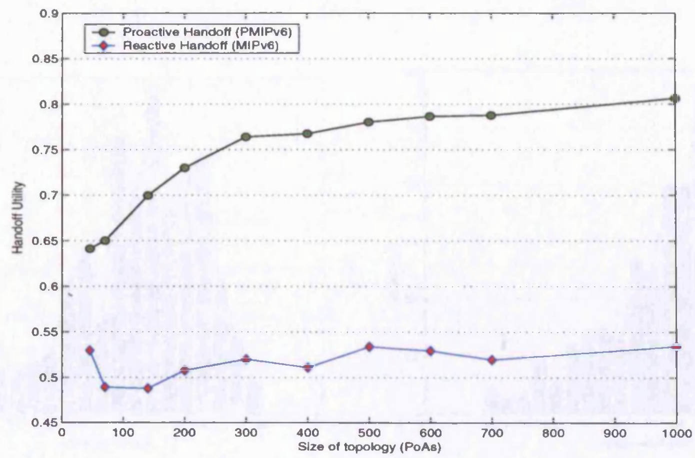
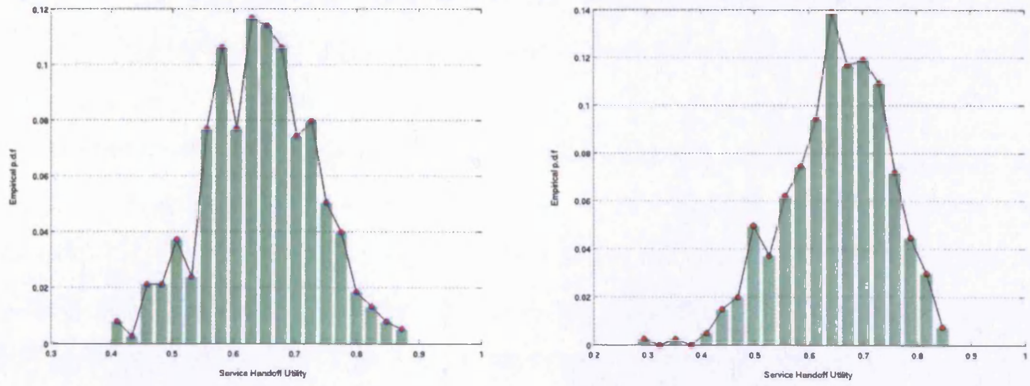


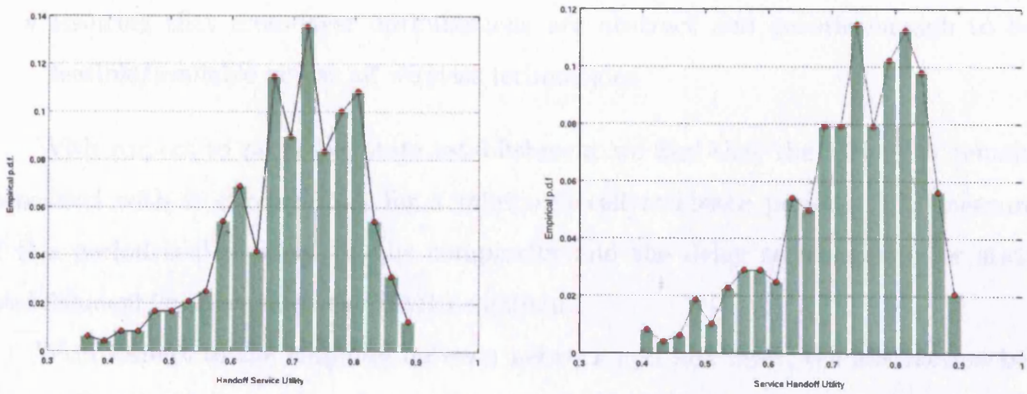
Figure 4.33: Measure of handoff utility as a function of the size of PoA topology within a geographical region.



(a) 45 ARs

(b) 70 ARs

Figure 4.34: Empirical probability density function of Proactive Handoff Utility experienced over 45 and 70 PoA topologies for 10 MN @ v=10m/s and p=10sec



(a) 140 ARs (b) 200 ARs

Figure 4.35: Empirical probability density function of Proactive Handoff Utility experienced over 140 and 200 PoA topologies for 10 MN @ v=10m/s and p=10sec

4.12 Conclusions

Proactive handoff management has been shown to eliminate reactive handoff delay arising at the network layer, by:

- identifying the immediate handoff AR neighbours with respect to MN's current point of attachment through Handoff AR discovery (HARD).
- promoting state establishment pertinent to MN's (IPv6) network connectivity at the next PoA, well in advance of its IPv6 handoff transition.
- establishing a sufficiently abstract mapping between the network and the link layer that allows expedient movement detection at the network layer with no reliance on network layer signalling (such as router advertisements)
- ensuring that cross-layer optimisations are abstract and generic enough to be feasible/available across all wireless technologies

With respect to proactive state establishment we find that the MN must remain associated with it current PoA for a minimum cell residence period. The measure of this period is dependent on the complexity and the delay requirements for state establishment/evaluation at the HAR neighbour.

With respect to the mapping between network and link layer, the abstraction between a Routing (network-layer) and a Mobility (link-layer) Neighbourhood enables the MN to transform physical node movement into mobility-hop Roaming state. Such transform effects a proactive IPv6 handoff with no dependence on traditional functions of IPv6 neighbour discovery such as Address Resolution or router advertisement signalling.

With respect to cross-layer optimisations proactive handoff management ensures that generic information available at the link layer are employed. While it has provided a much more elaborate set of link layer information that are essential for general proactive mobility management deliberations at the MN, the handoff management function has essentially required a single type of information: *AP identification information* indicating the Mobility Neighbourhood component of the associated PoA. Such type of information can be safely generalised for *any* wireless networking technology, since APs must remain at all times identifiable for management purposes.

In this manner, proactive IPv6 handoff management signalling reduces both the measure of required signalling and its associated probability of MAC contention. From this perspective, the measure of MAC-sublayer contention experienced under a proactive

IPv6 handoff is smaller than that of a reactive MIPv6 handoff, since it requires *no* router advertisements/solicitations or address resolution signalling for the purposes of address auto-configuration. Under a proactive IPv6 handoff, the measure of MAC contention appears to be essentially the same as the one experienced by an IP-stationary wireless host.

4.12.1 Handoff Delay Performance

From the performance analysis of proactive versus standard reactive handoff management we may conclude that proactivity handoff management *can* address successfully delay seamlessness at the *network* (IPv6) layer. Section 4.10.1 has shown that the proactive handoff management function introduces reduction in the measure of handoff delay by a factor of four compared to its reactive MIPv6 counterpart; in particular, under reactive MIPv6 the MN experiences a handoff delay in the order of 1.614-1.9 sec, whereas proactive MIPv6 incurs a delay between 420-480ms. With respect to jitter, a reactive MIPv6 handoff observed a delay variance of around 118ms while a proactive handoff experiences only about 30ms. With respect to the measure of packet loss, reactive MIPv6 handoffs experience packet loss runs around 88-90 packets, while during proactive handoffs the loss runs are reduced down to about 22 packets.

Focusing on handoff delay, the observed measure of proactive handoff delay is well above the 200ms requirement imposed by interactive real-time services. Thus, while proactive handoff management *can* eliminate delay incurred by network layer functions during an IPv6 handoff, it is found to be insufficient *by itself* to address the *total* handoff delay.

The measure of delay incurred in proactive handoffs is owed primarily to two latency factors over which the network layer can exert no control: (i) the link-layer handoff delay, (ii) round trip time delay. In the observed proactive handoff delay measure, 380-420ms of the delay is owed to the link-layer handoff, while 80-100ms to the average round trip time delay.

It is important to note that for the purposes of simulation the underlying wireless model implemented the measure of L2-handoff delay observed in a particular IEEE802.11 WLAN vendor implementation (Cisco 350) during the experimental study of Chapter 3. However, different WLAN implementations and in general different wireless technologies achieve different L2-handoff latencies [105, 280]. For instance, certain WLAN implementations can achieve L2-handoff delays as low as 150-160ms. On the contrary for cellular networks the L2-handoff delay may be as high as 1-2sec [280].

Assuming the figure of 200ms as the hard delay bound, it appears that provision of any realistic guarantees towards delay seamlessness in proactive IPv6 mobility management must impose a maximum delay bound on both end-to-end (CN-new PoA) and L2-handoff delay $\leq 150ms$ for highly interactive IP applications. In this manner, delay transparency may be addressed by ensuring that MN's traffic is also *redirected* towards its new PoA while a proactive IPv6 handoff is in progress. To this end, flow redirection/forwarding management emerges as an essential function in proactive MIPv6 management if delay seamlessness is to be supported.

Unless the above delay bounds can be guaranteed statistically on the measure of end-to-end delay between the CN and the new PoA, as well as the measure of L2-handoff latency,*no* delay seamlessness guarantees can be provided by proactive handoff management even if complemented by some form of flow forwarding management.

The latter observation holds not only for proactive MIPv6 management but *any* MIPv6 management mechanism that attempts to provide hard delay bounds ≤ 200 in handoff delay during interactive communications. Chapter 5 evaluates the viability of this hypothesis.

4.12.2 PoA divèrsity and Service Utility

With respect to the performance benefit as a result of handoff selectivity during MN's next IPv6 handoff, results have shown a significant increase in MN abstract service utility when the MN is enabled proactively with choice about its next PoA amongst a set of handoff candidates. In particular, MNs are achieving an immediate performance improvement of about 14% percent simply by means of availability of choice in PoA handoff selection for the nominal measure of 45 PoAs. The benefit increases rapidly to about 22% as the measure of PoA increases to about 300. From there and up until a topology of 1000 PoAs the measure of service utility increases only marginaly by 5%, yield a total performance benefit increase of about 30%. The reason for such a small increase between 300 and 1000 PoA is emerging to be the issue of false positive choices in PoA handoff selection. In particular, due to the high density of PoA the MN appears to chose the PoA with the highest service utility measure, which however is only peripheraly incident on the core direction of MN movement trajectory. As a result the MN is obliged sooner than expected into selecting a new PoA, as it moves out of the transmission range of the one previously selected.

From the analysis, it becomes clear that where proactively established context state (e.g. IP Roaming) is not available, reactive MIPv6 management must be available. In this light, Proactive and standard reactive MIPv6 management are found to operate best in complementary roles particularly where HAR neighbourhood information is yet either unavailable or unattainable.

Ultimately it is important to note that the principle of proactive state establishment, analysed in the context of IP Roaming state for the purposes of critical IPv6 connectivity with new PoAs can also be generalised for other types IP connectivity state such as AAA or QoS. This is possible, since the resolution of such context can be done much in the same manner (like IP Roaming state) during the beginning of MN's cell residence period at the current PoA. This provides tentative availability of AAA privileges or QoS reservation at HAR neighbours at the cost of increased signalling overheads. However, as we see in the Chapter 6 such additional signalling cost is compensating for MN non-determinism in its mobility pattern as well as frequent transient ping-pong effects frequently observed in wireless networks.

Chapter 5

Seamless Flow forwarding management using HandoffCast

5.1 Introduction

Thus far, results in Chapter 3 have shown that during MN's IPv6 handoff, its ongoing VoIP (or any) IPv6 flow gets severely disrupted by two collimating factors. The *first* pertains to the delay incurred due to reactive establishment of state pertinent to MN's IP connectivity (IP-Roaming) during its IPv6 handoff; the *second* pertains to the temporal *disruption* of packet transmissions towards the MN, during its IPv6 handoff.

To address these two factors Chapter 4 has presented a proactive mobility management architecture of which the handoff management aims to *reduce/eliminate* the delay arising during MN's IPv6 handoff, in support of interactive real-time application services.

A detailed evaluation of this architectural component showed that proactive handoff management can effectively eliminate any latency incurred by the *network* layer. This, however, fails to be the case for delay components arising at the link layer, or as a result of external factors such as (round-trip) delay due to congested paths. Proactive handoff management was found to exert no positive influence towards reduction of these latency components dominating the measure of *persistent* handoff delay. We define persistent handoff delay the measure of delay that *persists* during a proactive handoff as a result of factors[1] *beyond* the control of the network layer.

To alleviate the aforementioned limitation, investigations in Chapter 4 concluded that the measure of *end-to-end* delay between the CN and the new PoA, as well as the measure of *L2-handoff* delay for the wireless technology at hand, must stay below the

[1]measure of L2-handoff delay characteristic of the wireless technology and round trip time delay as a result of link congestion

threshold of $150ms$. This would enable subsequent mobility management extensions to enhance *feasibly* the proactive handoff delay performance.

5.1.1 Diversity of L2-handoff delay amongst wireless technologies

A brief investigation over different wireless technologies reveals, however, significant disparity in link-layer handoff delay performance. For wireless technologies with regulated frequency bands where channels are allocated explicitly per MN (e.g. GSM/GRPS/UMTS), hysteresis-based handoff techniques [283] allow soft handoffs with effective cellular link-layer handoff delays in the order of $< 50ms$, for average traffic load and MN power requirements [284, 285, 286]. Such L2-handoff delay performance is owed to: (i) strict frequency allocation per communicating terminal, (ii) the ability of wireless MN terminal to detect the pilot signal of *multiple* cellular Bases Stations (BS) *simultaneously*, (iii) maintain associations with multiple BS for handoff control purposes.

On the contrary for local (or metropolitan) area wireless technologies (e.g IEEE802.11b/g/a, 802.16), as evaluated in Chapter 3, L2-handoff latencies can range between 150-430ms, depending on the vendor implementation, for nominal traffic load and power requirements. This is primarily due to: (i) deregulation of the operating ISM band, allowing unstructured channel assignment per AP (ii) pilot signal detection and association with a *single* AP at any time (iii) scanning of the entire range of 'channels' available with the operational ISM band, since frequency reuse/allocation is not strictly enforced.

Thus, for mobility management purposes, the differentiating factor between the two broad classes of wireless technology, namely IEEE 802.11/16 or cellular, appears to be the magnitude of L2-handoff delay period; cellular technologies exhibit, albeit at a higher deployment/management cost, a significantly smaller L2-handoff delay than IEEE 802.11/16 technologies in their core specification.

It can be seen that, coupled with the average round trip time delay (85-100ms), cellular L2-handoff latency effects a measure of persistent handoff delay in the region of 130-150ms; it implies that temporal *disruption* of MN's IP flows becomes more pronounced for IEEE802.11/6, than cellular wireless technologies. The latter indicates that, 802.11/16 technologies require further optimisations at the link layer to achieve a manageable measure of L2-handoff delay, if proactive mobility management is to address successfully delay seamlessness.

5.2 Problem Description

Despite the differentiation in L2-handoff delay performance between cellular and IEEE 802.11/16, however, proactive handoff management remains insufficient, by itself, to address delay seamlessness during an IPv6 handoff, although it performs significantly better than current reactive MIPv6 management standards.

This is because, during this period, all (HA and CN) peers continue to transmit packet flows destined for the MN towards its *last* PoA, notwithstanding MN's departure to a new PoA. This temporal disruption of packet flow towards the MN persists until the latter ultimately updates its bindings with its HA and CN peers.

For connection-oriented (TCP) transport protocols, IPv6 flow disruption is experienced as *transport-level retransmissions* [287, 280]. In extreme cases of excessive one-way delay, TCP flows may be disrupted to the extent of connection resetting [43]. To this end, a number of solutions have been proposed by [287, 288, 289, 290, 291, 292], ensuring responsiveness of the TCP function in the face of wireless medium access, manifested also through L2-handoffs. For a detailed analysis of TCP transmission issues over wireless links with particular emphasis on GPRS and 3G Cellular, see [293].

On the contrary, connection-less transport protocols such as UDP/RTP which support multimedia application services, IPv6 flow disruption is experienced by the MN as *packet loss*; UDP, employs typically no transmission control mechanism to provide any kind of assurance in packet delivery. In addition, the lack of instant notification upon the handoff-departure of the MN from its current PoA, gives rise to a temporal *black hole* effect; packets sent towards the MN, reach a dead-end as they arrive at the last PoA, since the latter: (i) has no indication of the MN's departure, (ii) has no information about MN's new PoA, (iii) has no mechanism provision for redirecting outstanding MN traffic. As such, during an IPv6 handoff, in-transit packets are never received by the MN and thus, rendered *lost*, as shown in figure 5.1.

For (non-interactive) streaming application services, it is up to the application to enforce a streaming play-out buffer [294] coupled with its own retransmission mechanism so as to damp any disruption in the communicated IP application service [295]. For instance, in the case of the 2.6-3sec handoff delay experienced under Mobile IPv6, streaming application services can perform satisfactorily by enforcing a play-out buffer[2] of 2-3 sec at the receiver. A tightly-coupled application-level retransmission mecha-

[2]This buffer is generated at initialisation time of the streaming service and accompanied by a corresponding play-out delay

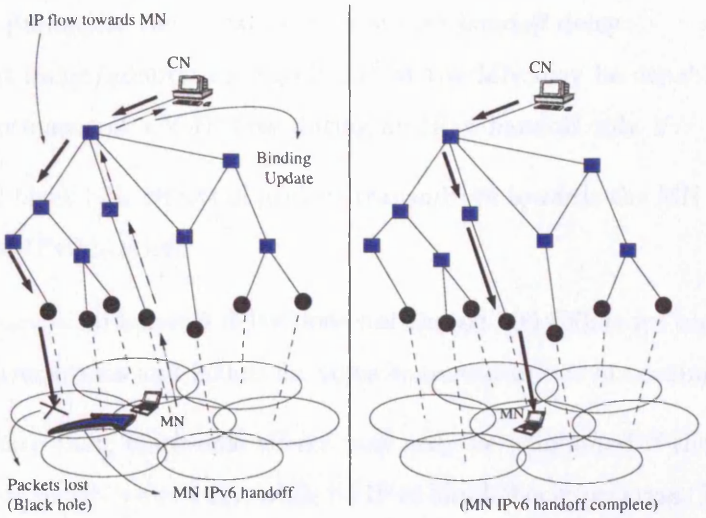


Figure 5.1: Temporal black hole effect during MN's IPv6 handoff under standard MIPv6

nism, then ensures that packets lost during an IPv6 handoff are retransmitted by the sender, while the receiver consumes packets filling its play-out buffer. In this manner, non-interactive streaming services can enjoy acceptable performance over reactive (or reduced proactive) IPv6 handoff delays.

For interactive real-time services such as VoIP, a large play-out buffer delay is not viable for a number of reasons: (i) human interactivity requires a one-way delay bound of 100-150ms if communication is to sustain intelligibility [296], (ii) packet delay beyond the aforementioned bounds translates to *late* loss rates, since the play-out schedule of severely delayed packets impacts negatively intelligibility and thus, the flow of communication between humans, degrading its subjective service quality [297, 217].

Looking at wireline or stationary wireless networks, a voice call is reported to tolerate up to 10% packet loss rates according to [36]. Such loss rates are *perceptibly* tolerable (no annoyance) with the aid of receiver loss replenishment techniques such as packet loss concealment (PLC) [298]. These techniques rely on signal interpolation which in turn perform well under short packet loss runs (1-2 packets) [204]. Packet loss runs, however, during an IPv6 handoff are significantly longer (15-40 packets), for a nominal packetization rate of 20ms, than that managed by PLC techniques. Such a characteristic, precludes PLC as a viable technique of packet loss recovery during an IPv6 handoff.

From the above it may be seen that, stringent delay bounds for interactive real-time services make both packet retransmissions or packet loss concealment techniques impractical to pursue for the measure of persistent handoff delay.

A play-out buffer/adaptation algorithm[3] at the MN may be capable of enhancing the delay performance of a VoIP flow during an IPv6 handoff only *if*:

- temporal black-hole effects in packets transmitted towards the MN are *eliminated*, during an IPv6 handoff.
- the measure of end-to-end delay does *not* exceed 100-150ms for highly interactive voice conversations and 200ms for voice communications of nominal interactivity.

It is intuitive that, black-hole effects may only be eliminated if the previous PoA *forwards* packets to MN's new PoA, while its IPv6 handoff is in progress. With respect to the measure of end-to-end delay such guarantee can only provided if: (i) the L2-handoff delay of the underlying wireless technology remains below 150ms and (ii) the 95% percentile of the one-way delay between the CN and the new PoA can be guarranteed to remain below 150ms.

Annex F.2 presents a brief assessments on the empirical measure of handoff rate drawn from cellular voice communications.

5.2.1 Hypothesis

The measure and rate of IP flow disruption emerging during an IPv6 handoff, calls for additional mobility management functions to leverage its effect over interactive real-time services. Such function appears to be the one of flow forwarding between the previous and new PoA while MN's IPv6 handoff is in progress.

We, thus, argue that from the perspective of the mobility management function, *it is essential that packets transmitted towards the MN, pursue the MN towards its new PoA, instead of being lost at the previous PoA.* In this manner, black-hole effects can be eliminated, while the delay bound for the packet flow *sustaining* transmission towards the MN, is tracked by the one way delay between previous and new PoA.

This chapter investigates the effects of a multicast-based, proactive flow forwarding management mechanism, for the purposes of: (a) eliminating black-hole effects due to handoff, (b) leveraging the measure of persistent handoff delay during MN's IPv6 handoff. In this manner, the overall mobility management architecture can address delay performance beyond the control of proactive handoff management at the network layer. We argue that *proactive flow forwarding with minimal buffering at handoff PoA candidates, is capable of providing delay seamlessness over wireless networks towards support*

[3] Annex F.1 provides a short elaboration on play-out adaptation algorithms.

of interactive real-time services. Such techniques are capable of reducing significantly or eliminating the associated packets loss, without resorting to application level packet loss management techniques such as forward error correction (FEC), packet redundancy or interleaving [204].

Furthermore, the dependence of the measure of persistent handoff delay on L2-handoff delay performance and the identification of more pronounced flow disruption effects in terms of delay and associated packet loss over IEEE 802.11 technologies in particular, calls for significant reductions on the respective measure of L2-handoff delay. To this end, we argue that *proactive mobility management can provide sufficient state in a forward manner, to support successfully such L2-handoff delay reductions over IEEE 802.11/16 wireless technologies.* The emerging principles from the proposed mechanisms/functions, in support of such hypothesis, remain generic enough to effect similar L2-handoff delay optimisations over other wireless technologies, in need of such support, for mobility management purposes.

To this effect, this chapter investigates further a minimal set of cross-layer optimisations between the wireless link-layer manifested through the AP subsystem of a PoA and the network layer represented through the AR device, as an integral part of MN's current PoA.

5.2.2 Outline

Section 5.3 presents work related with flow forwarding for the purposes of mobility management. It describes further existing work in L2-handoff delay performance and existing optimisations to reduce the measure of link-layer handoffs over modern 802.11b/g WLAN systems.

Section 5.4 presents the system design of HandoffCast, a multicast-based flow forwarding management function in support of delay seamlessness, by guarding proactively against delay factors emerging below the network layer.

Section 5.5 evaluates the performance of proactive IPv6 flow forwarding management. Subsequently, Section 5.6, evaluates HandoffCast performance with respect to interactive real-time delay guarantees, through multicast-based flow forwarding.

Section 5.7 presents a summary of our findings concluding on the performance of HandoffCast as a viable flow forwarding function, extending the overall proactive IPv6 mobility management architecture.

Additionally, Annex F.4 presents the set of generic cross-layer optimisations proposed, in support of expedited IPv6 mobility management through low-delay link-layer handoffs over IEEE 802.11 networks. The focus of these optimisations remains on 802.11 wireless LANs. Cellular networks achieve significantly lower L2-handoff delays and thus, are not expected to affect delay guarantees supported by proactive IPv6 mobility management (assuming native IPv6 signalling).

Annex F.5 presents an assessment of a proposed 802.11 link-layer (L2) handoff optimisation, targeting low L2-handoff delays.

5.3 Related Work

5.3.1 Flow forwarding

Malki and Soliman propose a flow forwarding mechanism identified as bi-casting [299]. This mechanism encompasses signalling from the MN towards the CN with CoA information about MN's candidate new PoA. On receipt of such signal the CN unicasts temporally the communicated packet flow towards the MN to two CoA destinations: the current and the candidate CoA address. While such approach does *not* require alteration at the PoAs themselves, it introduces packet flow duplicates[4] across the entire path between the CN and MN at two candidate PoAs. The effect is exacerbated *end-to-end* when the communicated IP flow is duplicated to more than two (n-cast) candidate MN CoA destinations. This implies unnecessary loading of the Internet backbone with redundant packet transmission and thus, inefficient use of bandwidth.

Perhaps the most popular approach of *smoothing* MN's IPv6 handoff, that is, minimise packet flow disruption during an IPv6 handoff, is the one of unicast tunnel forwarding from the previous to the new PoA. This was originally proposed in Mobile IPv4 [31] where routers acting as foreign agents could also act temporarily as intermediate home agents for the purposes of packet interception and forwarding (to the new PoA) until MN's handoff was complete. Such capability remained optional, while being redressed in a later working draft specification of the Low latency handoffs proposal [300].

Unicast tunnel forwarding has been proposed by a number of evolving [5] IPv6 mobility management protocol specifications [300, 301, 33] aiming to support smooth IPv6 handoffs.

In the form employed by these mobility management protocols, PoA tunnel for-

[4]at both wireline and wireless segments.

[5]These draft IPv6 mobility management recommendations are the subject of on-going standardisation effort. As a result their specification is subject to continuous changes, making it very difficult to assess the validity of any performance claims made. For this reason our assessment focuses on core ideas that have remained reasonably constant during the cycle of draft revisions.

warding or bi-casting, enables unicast tunnelling of packets arriving at MN's previous AR to a *single* new HAR candidate. Such form of *path re-routing* requires, however, that the new HAR candidate is identified in advance of MN's IPv6 handoff.

The unicast tunnel arrangement of the aforementioned handoff management schemes assumes, almost invariably, an L2-handoff trigger among available new APs to determine precisely the *exact* new PoA. Such assumption, however, is not reflected by modern cellular systems. In such systems, a soft handoff is attainable by: (i) the ability of the MN to detect pilot signals from multiple BSs, (ii) the ability to remain connected with an active set of (neighbouring) BSs at any given time [302] while in soft handoff mode (iii) receive traffic from multiple BSs of the active set while in soft handoff mode.

These three factors imply that cellular technologies do not provide information about the exact new PoA. Instead, they rely on signal diversity, whereby the MN combines received signals from multiple BS to sustain communication quality during a cell handoff.

With respect to IEEE 802.11 technologies, while such assumptions may be valid for WLAN hosts operating in ad-hoc mode, they do not hold for WLANs operating in infrastructure mode; in such mode, 802.11b link-layer handoffs are by-design reactive. In infrastructure mode, a WLAN MN may be associated with *only one* AP at any time. Hence, the MN *cannot* receive signal strength information from, and thus identify, an active set of neighbouring PoAs (i.e. their AP subsystem) even if it exploits the channel leakage of its spreading over the operating channel.

Detecting APs incident to MN's transit path, within its M-neighbourhood, requires at least *one* change in MN's operating frequency channel followed by probes requests (see Chapter 3); such scanning phase requires that the MN detaches from its current PoA, initiating thus an L2-handoff. It is reminded that in the vast majority of WLAN 802.11b/g implementations an L2-handoff implies scanning of all 13[6] 'channels'.

Each of these 'channels'[7] require a minimum dwell time for probe purposes which varies according to the vendor implementation. The longest dwell time has been observed to be in the region of 38ms [303]; such figure is validated also by results of Section D.6.1. A quick estimate reveals a probe delay period of at least 380ms per scanning attempt for the slowest vendor implementation. Even by halving such dwell time, the

[6] 11 for North America under FCC.
[7] more accurately frequency sub-bands

probe delay alone approaches the hard delay bounds of interactive real-time services, without accounting for round-trip delays. Such delay is, clearly, disruptive for MN's on-going packet communication with its peers. Annex F.3 presents related work on mechanisms aiming to reduce link-layer handoff delay in IEEE 802.11 wireless LANs.

As a result, fast handoff proposals that rely on the ability to anticipate an imminent link-layer handoff by means of comparing signal strength between AP neighbours cannot be applied to IEEE 802.11 WLANs operating in infrastructure mode. It follows further that, since the MNs cannot receive MAC frames transmitted from other APs within its M-neighbourhood, it is *not* possible for the MN to receive multiple router advertisements from neighbouring ARs and maintain a list of neighbouring PoAs *a priori* for future use.

It may be seen that, for the purposes of either unicast PoA tunnel forwarding or bi-casting, prediction of the **exact** next AR in advance of MN's handoff, is expected to be highly *inaccurate* in a multi-cell WLAN M-neighbourhood due to: (i) the short-term non-determinism in the mobility pattern of the MN (ii) the availability of multi-AP overlap. While link-layer signal strength hints *may* suggest that the MN moves *away* from the the current PoA, they *cannot* indicate *the direction* in which such movement is made. Moreover, even by means of interleaved detection of multiple APs, the availability of multiple beacons does not provide sufficient discriminators about MN direction to resolve accurately the handoff PoA candidate. Hence, link-layer signal strength, *does not* necessarily provide a meaningful hint about MN's next AR handoff candidate in the majority of managed handoffs.

The above imply that over 802.11 WLAN links, in-advance path re-routing under unicast tunnel forwarding or bi-cast meets significant limitations, since the exact new PoA cannot be determined in advance over WLAN networks, either through L2-handoff triggers or signal strength hints. To this end, we argue that mobility management proposals, such as Fast Handoffs [33] or low latency handoffs [300] which predominantly use unicast tunnel forwarding or bi-casting to effect a smooth handoff, base their claim on idealised mobility scenarios that do not reflect handoffs under multiple PoA handoff candidates or ping-pong effects. Such effects occur with significant frequency in cellular environments [304]. For a more detailed discussion on ping-pong effects see Section 5.4.8.

5.4 HandoffCast: Proactive flow forwarding during IPv6 handoffs

Focusing on flow forwarding management, we have seen that proactive path rerouting through bi-cast fails due to unpredictability of MN's direction of movement: it cannot identify accurately a single PoA candidate in advance of MN's next IPv6 handoff for the majority (e.g. 90-percentile) of MN's handoffs.

Extending bi-casting to tunnel packets to multiple new PoA candidates is essentially a very inefficient form of *multicast*, since it requires k copies of the same packet to be unicast to all (k) PoA members of the R-neighbourhood. On the contrary, a *multicast-based* path rerouting mechanism employed during an IPv6 handoff, removes such inefficiency, since it forwards a single copy of the packet to AR handoff candidates, members of the R-neighbourhood.

Multicast-based rerouting has been applied so far either at the *edge* of a network domain in micro-mobility proposals [232, 231], or *end-to-end* in certain macro-mobility proposals [184], at the HA/CN peers towards the MN. Each approach balances certain trade-offs; initiating multicast forwarding at the edge of the network carries the complexities of single point of failure, and traffic concentration over a single link, traded-off for optimal path re-routing. Supporting multicast forwarding end-to-end, generates multicast trees that are severely limited in terms of scalability, as they span multiple different administrative domains between the source and the receiver. It is reminded, that one of the fundamental limitations in the deployment of multicast has been the lack of control on cross-domain routers participating on the construction of the multicast delivery tree.

To combine positive benefits from multicast forwarding, while leveraging unwanted trade-offs from the above approaches, this investigation looks at supporting multicast forwarding at MN's *previous* point of attachment (PoA). To this end, we augment the proposed mobility management architecture with a tightly-coupled, *multicast*-based, path re-routing mechanism, that sustains packet flow towards the MN through PoA neighbours, during MN's IPv6 handoff; such mechanism is collectively identified as *HandoffCast*.

HandoffCast is a multicast-routing protocol abstraction, that redirects transmissions destined for the MN, by having IPv6 flows *pursue* dynamically MN's mobility pattern during a handoff, within its current M-neighbourhood. Disruption of communication flows can thus be avoided, while packet loss during MN's handoff is minimised, at the extent allowed by the residual measure of persistent handoff delay.

Since bi-casting is a special case of multicast, it is intuitive that a multicast-based tunnel forwarding mechanism, performs as well, during MN's IPv6 handoff. HandoffCast, makes further provisions in allowing a non-deterministic MN mobility pattern to bounce freely over candidate PoA while engaged actively in interactive packet services.

5.4.1 Architectural Overview

An IPv6 CoA identifies state for two functions essential for the reachability of a host behind a visited network: *addressing* and *routing*. For each change in MN's CoA, its addressing and routing will also change. Such change becomes imperative during MN's IPv6 handoff, to sustain packet transmission with its peers. HandoffCast takes on the view that *for the duration of the handoff, the MN must refrain from changing its addressing and routing*, that is to say, the MN must refrain from changing its CoA *during* an IP handoff.

Given that a unicast CoA address becomes topologically *incorrect* as soon as the MN changes its subnet PoA, it is becomes unrealistic to require that the MN avoids the explicit change of its unicast CoA during a handoff. Instead, by employing multicast, the HandoffCast forwarding function ensures that MN's addressing and routing remains intact for the duration of its handoff.

To achieve minimal management costs, it is essential that the associated HandoffCast identifier *stays the same* at least within a single administrative domain, independent of the visited PoA, for the duration of the IPv6 handoff. This can be achieved through the *multicast* group CoA identifier. For the purposes of HandoffCast such identifier is referred to as *Handoff Care of Address (HCoA)*.

In IP-Multicast routers 'conspire' to abstract routing and addressing behind a single, network-independent identifier for the purposes of efficient, scalable multi-receiver source transmissions. In a similar manner, HandoffCast abstracts further the underlying multicast routing function, in the context of handoff management; by 'inviting' routers to support path rerouting of packet flows towards the MN over its HAR candidates, HandoffCast targets to eliminate flow disruption during MN's handoff. Such goal is in turn dependent on a persistent handoff delay bound $< 200ms$.

To minimise the disruption of IPv6 flow transmissions towards the MN, *HandoffCast* requires that IP Roaming state has been established proactively at the current

PoA[8] (AR_c) ahead of MN's next IPv6 handoff, as presented in Section 4.8.

Once established, components of IP-Roaming state (valid within MN's R-neighbourhood) are mapped, through HandoffCast, onto MN's allocated HCoA, uniquely identifying the MN during the period of its handoff. The mapping allows CoAs proactively allocated at the HAR neighbours, to *join* the HCoA address, ahead of MN's IPv6 handoff.

It is important to note that HandoffCast forwarding management aims to address delay-bound path rerouting for MN's on-going IP flows, in a robust manner; that is, independent of MN's non-determinism in its mobility pattern. Given that exact determination of MN's next PoA, in advance of MN's handoff is frequently infeasible, due to propagation or ping-pong effects, a robust solution would be required to consider *multiple* PoA handoff candidates and effect multiple signalling interactions in preparing MN's next IPv6 handoff. On these grounds it may be seen that robustness is unavoidably traded-off with increased signalling per mobile node.

It is thus, essential to acknowledge that HandoffCast adopts this trade-off in favour of robust performance targeting delay seamlessness guarantees; that is to say, HandoffCast is expected to support delay seamlessness during flow forwarding, at the cost of increased signalling for the purposes of HCoA group management, on a per MN basis. The emerging question is what is the measure of signalling overheads on an MN basis? under what conditions can such overheads hinder the scalability of the proposed proactive mobility management optimisation? These issues are addressed at the performance evaluation of HandoffCast in Section 5.6.

Upon detection of MN's imminent handoff, HandoffCast is triggered to initiate flow forwarding, by having AR_c re-route MN's traffic towards MN's HCoA for the duration of the handoff. Multicast routing undertakes the dispatch of packets to candidate PoAs with ultimate destination MN's tentative CoAs. Hence, traffic forwarded towards MN's HCoA, continues to pursue MN's ultimate (unicast) CoA destination.

Packets arriving at a PoA neighbour are first *buffered* and conditionally forwarded over the wireless link, *if* the MN appears on that link. A *circular* buffer of 200ms worth of packets is adopted for each of MN's flows. While the MN remains absent for more than 200ms the PoA neighbour drops packets at the head of the buffer as new packets populate its tail. Such buffering strategy at the AR neighbour is followed because: (i)

[8]For the purposes of HandoffCast management the terms PoA and AR are deemed equivalent and hence used interchangeable for the rest of this chapter.

holding packets longer than 200ms for interactive real-time applications yields a *late* loss rate; (ii) the MN has appeared on the link of a different AR neighbour and thus, has already received a copy of this traffic.

The MN is assigned an HCoA once, when it first enters a new WISP domain. To this end, HandoffCast assumes that multicast deployment is available on an *intra-domain* basis. Allocation of the multicast CoA per MN is performed by means of a simple mapping between the allocated unicast CoAs at PoA neighbours and MN's HCoA. It is noted that since the scope of the multicast CoA is domain-wide, the *same* HCoA may be used by the same MN at a new domain, as long as this is not used by another MN in that domain. By adopting a strict algorithmic mapping between proactively allocated unicast CoAs for a single MN and its respective HCoA, the possibility of duplicate HCoAs between MNs across different domains can be eliminated.

5.4.2 HandoffCast Management Triggers

Two critical aspects in the performance of HandoffCast are: (i) the moment that flow forwarding is initiated from the current PoA towards its neighbours, (ii) when buffered traffic at a PoA candidate should be forwarded over the wireless link. These aspects are dependent on three distinct events potentially experienced by the MN during its IPv6 handoff: (i) detachment (ii) attachment (iii) ping-pong. Ping pong effects represent about 15-22% of handoffs in cellular systems [305]. For 802.11 systems ping-pong effects are reported even for stationary MNs [306].

The best method of indicating any of these three event while eliminating the possibility of *false positives*[9] is by adopting a cross-layer optimisation through *link-layer (L2) triggers*. This is because, during an IPv6 handoff, the MN is subject first to a link-layer (L2) handoff *before* it effects a network-layer (L3) handoff, irrespective of the wireless technology.

The above argues in favour of the emerging fact that the overall IPv6 handoff process may only be expedited at the cost of *sacrificing* layer-independence, in a controlled manner.

Focusing on WLAN networks, such L2-trigger can be provided by means of the **de-association/ re-association** signal of the 802.11 link layer. This kind of notification information is readily available at both the AP and the MN through their respective 802.11 management function [307].

[9] A false positive with respect to handoff, is an apparently truthful indication of an imminent L2-handoff (precursor of an IPv6 handoff) due to SNR fluctuations which, however, does not give rise to an actual L2-handoff.

The *de-association* signal is exploited when the MN experiences either a *detachment* or a *ping-pong* from its current PoA. On the contrary, the *re-association* signal is exploited when the MN experiences an *attachment* to a new PoA. Figure 5.2 illustrates the three events and the use of association[10] signalling.

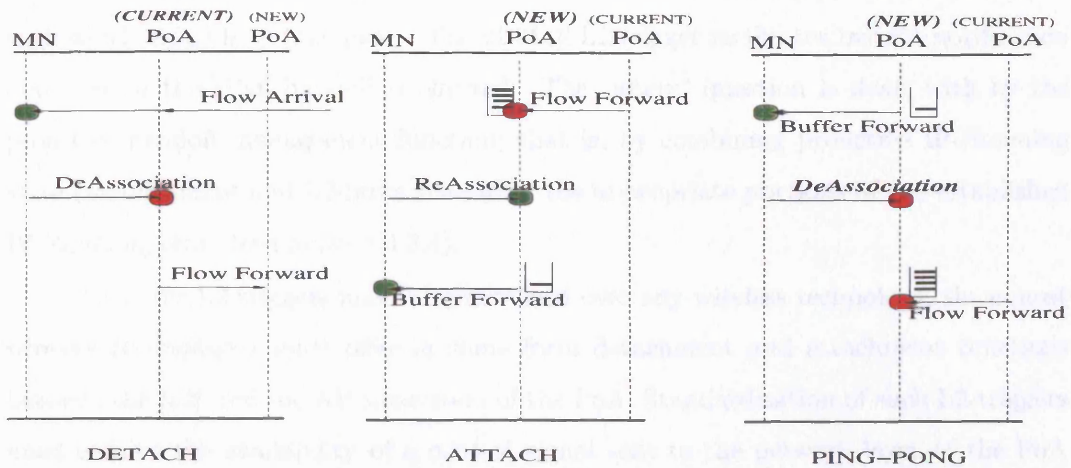


Figure 5.2: Events indicating an L2-handoff and use of de-association/re-association as reliable network-layer handoff triggers

In particular, when the MN detaches from its current PoA it may signal a de-association message. Alternatively the PoA (i.e. the AP subsystem) times out the particular MN association, when the number of link-layer retransmissions[11] exceed a threshold of 10 attempts. Under such configuration, link-layer retransmissions due to collisions must always be set at a significantly lower bound (about 3). This ensure that the AP can discriminate between collisions and de-association.

HandoffCast is initiated at the current PoA through the `DETACH` link-layer trigger. Buffer forwarding at the new PoA is initiated through the `ATTACH` link-layer trigger. In the event that the MN is detached from its *new* PoA due to either signal or movement trajectory fluctuations, the initiated HandoffCast is sustained until the MN settles at some new PoA. In such case, the `PING-PONG` trigger is enforced at the new PoA.

In this manner, the HandoffCast function sustains receipt of MN's traffic for the duration of the handoff, at the new PoA, *with no* dependence on the dispatch of Binding Updates (BU) towards MN's peers. The new PoA can be any HAR neighbour within the respective R-neighbourhood, supporting MN's movement pattern at 360^o degrees within the corresponding M-neighbourhood.

[10]In cellular systems the association signal is defined as *attach* signal.

[11]L2 retransmissions are effected on a μsec basis

On completion of MN's handoff, HandoffCast identifies the PoA neighbours for which availability of configured CoAs is not required anymore. Consequently, these PoA neighbours are requested to *leave* MN's HCoA address, since they do not belong to MN's *new* R-neighbourhood.

It is important to distinguish between *when* a handoff is to happen and *where*, (i.e. with which AR_n) is to take place. The `DETACH` L2-trigger facilitates instant notification about *when* the IPv6 handoff is effected. The *'where'* question is dealt with by the proactive handoff management function; that is, by combining proactive IP-Roaming state establishment and L2-hints activating the appropriate portions of pre-established IP Roaming state (see Section 4.8.4).

The above L2 triggers may be generalised over any wireless technology, since most wireless technologies must offer in some form detachment and attachment functions between the MN and the AP subsystem of the PoA. Standardisation of such L2-triggers must enforce the availability of a control signal sent to the network layer of the PoA system, when a wireless station is either detached from or attached to its AP subsystem.

5.4.3 Managing Handoff Care-of Addressing

A soft CoA (sCoA) tuple representing MN's tentative CoAs, is generated during state establishment by the handoff management function as elaborated in Chapter 4. Such a sCoA tuple is subsequently mapped to MN's HCoA address as shown in figure 5.3. The mapping is achieved through multicast group membership over the respective links of the PoA candidates.

It is noted that the mapping between the sCoA tuple and the HCoA address remains transparent to MN's peers. Both HA and CNs send packets towards the MN through AR_c with no knowledge about the existence of an HCoA address. During MN's handoff the previous PoA simply ensures that MN's traffic can be received over *any* PoA neighbour that has joined MN's HCoA address.

HCoA address allocation takes place only once at the first AR (home or visited) of every network domain accommodating the MN. The scope of the HCoA address remains domain-wide [308]. By bridging domain-wide multicast routing through inter-domain unicast tunnels, the MNs' HCoA address can attain *virtual* inter-domain scope and thus, support IPv6 handoffs also across domains. We postpone inter-domain HandoffCast management until later sections.

By bridging the multicast-enabled WISP domain-'islands' through unicast tunnels,

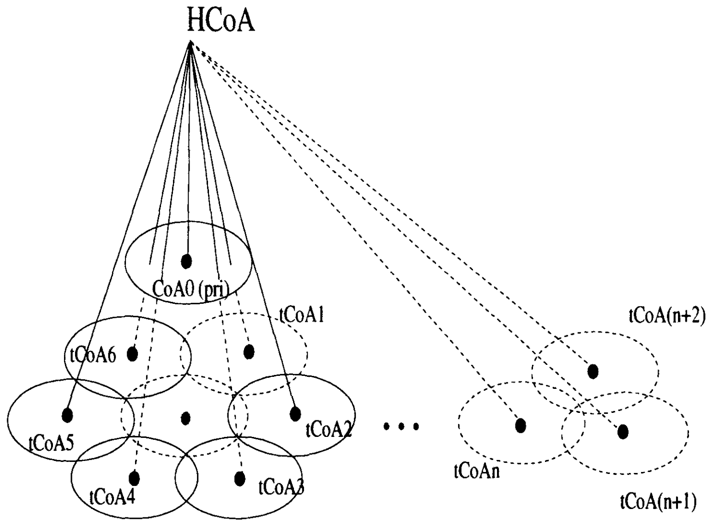


Figure 5.3: Mapping the HCoA multicast address onto unicast IPv6 tentative CoA unicast listener instantiations of the MN

the Proactive IPv6 mobility management architecture can support deployment over existing Internet infrastructures. However, such approach assumes a peering relationship between WISPs, in a manner similar to the CADENUS SLA framework [309, 310]. Similar peering relationships are currently in effect in cellular systems, and thus realistic to envisage. This is because the set of participating WISP has a joint interest to accommodate a portion of traffic, generated by mobile users, each over its own network, while sustaining access ubiquity.

Clearly, in the event that multicast IPv6 routing is supported also on an inter-domain basis, the scope of the multicast HCoA can be relaxed, with the HCoA identifier being valid across the Internet. In such event, the MN is allocated a permanent HCoA address, allowing for global handoff reachability.

5.4.4 HCoA address allocation algorithm

To allocate a HCoA address, HandoffCast devises a simple mapping between the MN's unicast CoA address and the generic multicast IPv6 address formats specified in [308]. Such a mapping is similar to unicast prefix-based multicast addresses proposed by Thaler et al. [311]. The need for a simple transparent multicast HCoA allocation mechanism is collision avoidance [312]. It is essential that the HCoA identifies a unique MN within the domain for the purposes of an IPv6 handoff; otherwise traffic destined for one MN would be erroneously be received also by another.

A unicast IPv6 address whether static or Care-of (i.e. mobile), must normally comply to an aggregatable format [313]. Under such format, 64 bits are allocated to the network identifier, while the remaining 64 bits are allocated to the interface identifier of the MN. The 64-bit *network* identifier is normally configured by the network provider at the ARs and characterises the network prefix of the network visited by the MN. It comprises of a 3-bit format prefix (FP), a top-level aggregation identifier (TLA_{ID}), an 8-bit reserved field, a 24-bit next level aggregation identifier (NLA_{ID}) and a 16-bit site level identifier (SLA_{ID}) as shown in figure 5.4. The NLA_{ID} allows the breakdown of an administrative domain into multiple provisioning sub-domain (ISPs). The SLA_{ID} enables each provisioning domain to identify its own hierarchy of subnets. It can be seen that the 64-bit network identifier changes for every subnet providing access to the MN.

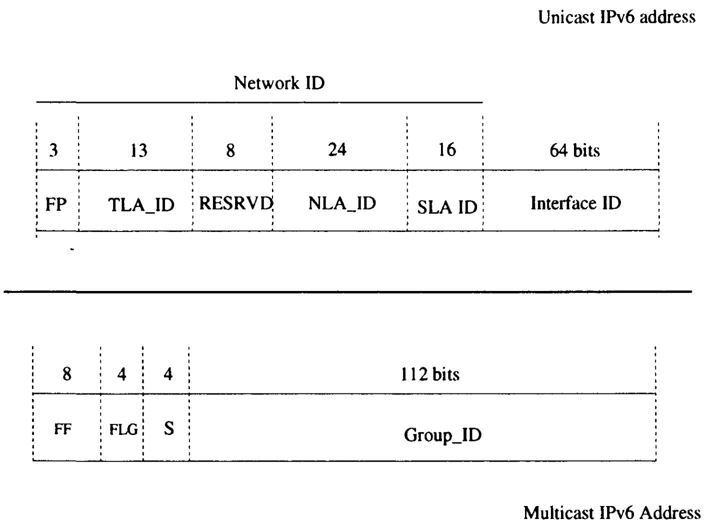


Figure 5.4: Comparison between standard unicast and multicast IPv6 address formats

The 64-bit *interface* identifier (ID) of an IPv6 address is typically generated[12] by means of a unique EUI-64 rule [314], acting on the link-layer identifier (aka MAC address) of the MN's wireless network interface card (NIC). The NIC's identifier may be described either by a fixed 64-bit [314], or 48-bit MAC address [109]. Irrespective of the MAC address size, the EUI-64 rule gives rise to the *same*, unique, interface identifier for the purposes of IPv6 address generation.

Alternatively, the interface ID of the MN may be described cryptographically, by means of its public key, by means similar to [281]. Such public key is associated with MN's wireless network interface and remains fixed to that interface. To ensure support

[12] other algorithms that yield a unique interface ID may be employed.

of multi-homed MNs, it is required that each *NIC* maintains its *own* public/private key.

In either of the two cases, the 64-bit interface identifier of the MN remains always the same for the IPv6 CoA to be obtained in each visited network, irrespective of the provisioning domain.

In contrast with unicast, a multicast IPv6 address, maintains a simple network-independent format as shown in figure 5.4. A prefix format (FF) indicating a multicast address, a 4-bit flag and a 4-bit scope for the multicast address. The remaining 112 bits are allocated to the group identifier.

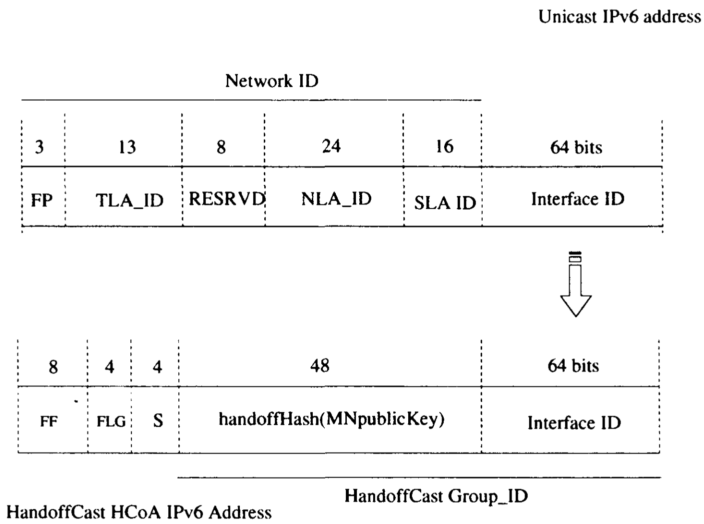


Figure 5.5: HandofCast-specific mapping MN's unicast address space onto its HCoA mutlicast address

Given the commonality of the interface identifier over MN's any IPv6 CoA, it is possible, for the purposes of HandoffCast management, to *personalise* a multicast HCoA address for the MN, with its unicast-based 64-bit interface identifier; this is sent to AR_c during IP Roaming state establishment. In addition, the MN can sign its 64-bit interface identifier with its private key, generating a 48-bit unique *handoff-hash.* This is shown in figure 5.5. In this manner the 64-bit identifier certifies a unique HCoA address bound to the particular MN during its handoff.

The purpose of such handoff-hash is two-fold: (i) it can be used for the purposes of non-repudiation during IP-Roaming state establishment when accompanied by a corresponding public key[13] for that MN, (ii) it can be used to increase the number of bits comprising the HandoffCast group identifier and thus reduce significantly the

[13]The strength of both the private and public key for handoff purposes need only be 48-bits, and be regenerated every few handoffs

probability of clashes.

The probability of an HCoA address allocation clash can be estimated combinatorially as follows. Let ρ be the number of available bits to randomise within a Handoff Care-of address in HandoffCast, and k the number of allocated HCoA addresses. Then $n = 2^{\rho}$ is the set of possible HCoA addresses generated by ρ bits. Combinatorially, the number of ways that k allocated out of n available HCoA addresses can be formed is given by the fundamental formula $n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$. For each of these equiprobable ways, the size of the population is $n_1 = n_2 = \cdots = n_k = n$ implying the size of the total HCoA address population is n^k. Hence, the probability of no clash at the kth IPv6 address is:

$$\begin{aligned} P(HCoA_k \neq HCoA_{k-i}) &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \\ &= \frac{n!}{(n-k)!\; n^k} \end{aligned} \tag{5.1}$$

Assigning $P(HCoA_k \neq HCoA_{k-i}) = q$, the probability of two multicast IPv6 addresses clashing on the kth generated HCoA address is:

$$P(HCoA_k = HCoA_{k-i}) = 1 - \frac{n!}{(n-k)!\; n^k} \tag{5.2}$$

Using Sterling's approximation we compute the probability of collision for 32, 64 and 112 bits. Figure 5.6 shows the probability of collision for the three discrete bit sizes.

As expected the size of HCoA group identifier for 112 bits, yields a much lower collision probability for a significantly large number of HCoA addresses; 32-bit identifiers yield 4.2950e+09 HCoA addresses before the number of collisions becomes prohibitive; a 64-bit HCoA group ID yields 1.84e+19 HCoA addresses, while a 112-bit identifier yield 5.19e+33 addresses before experiencing significant number of collisions in HCoA address allocation.

5.4.5 Intra/Inter-domain HandoffCast routing

HandoffCast is a sparse mode routing protocol that builds a single delivery tree per HCoA address, shared by all senders of the group. As such, its core routing function is based on the notion of shared trees *centered* at a multicast core through protocols

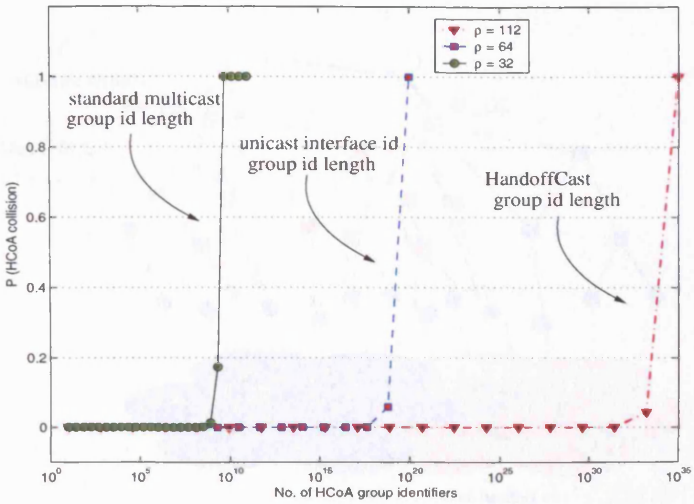


Figure 5.6: Probability density function of collision between allocated HCoA addresses, as a measure of varying group identifier size.

such as, bi-directional (sparse mode) protocol-independent multicast (PIM-SM) [315] or core-based trees (CBT) [316, 186].

To avoid traffic concentration on links near a single RP-router, HandoffCast employs the approach of a *distributed RP* set within each domain. To distribute the traffic load more evenly and thus, scale better across the network topology of a domain, each RP handles a specific *range* of HCoA addresses. This avoids further rebuilding the entire tree for each MN handoff, at the cost of suboptimal (triangular) routing [317]. Such design choice trades route optimality for protocol resiliency; this is because RP placement at the domain border with the Internet backbone, as followed by hierarchical schemes such as M&M or IDMP introduces single points of failure, while they concentrate all traffic over single nodes as seen in Chapter 2.

The suboptimal route trade-off does not impact negatively the performance of the protocol mechanism, considering that one-way delay over intra-domain paths is normally fairly small (with average 15-20ms with heavy tails at 30-40ms); thus, even by suboptimal (intra-domain) routes the one-way delay between previous and new PoAs is not expected to affect the delay performance of the communicated flow between the CN and the MN.

Intra-domain RPs may discover each other's identity by means of a dynamic bootstrap protocol [318] or, alternatively, given a small set, may be configured manually on PoAs. Placement of RPs is typically effected either by means of administrative selection or simple heuristics [319].

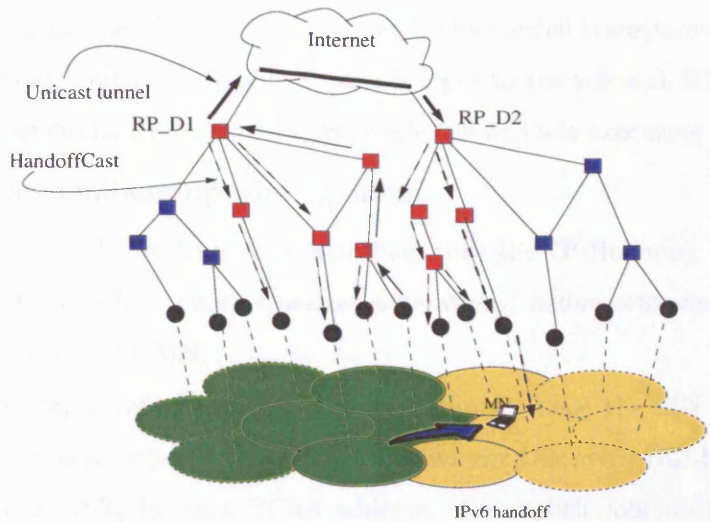


Figure 5.7: Inter-domain IPv6 handoff: bridging two RPs for bordering domains within the virtual R-neighbourhood of MN

HandoffCast remains functional across domains by *bridging* selected RPs between domains as shown in figure 5.7. To avoid multicast group state information accumulating at the border routers, HandoffCast employs a *peer-to-peer* routing approach typically found in content-sharing systems [320]. A minimal set of selected RPs from each domain are introduced through a peering relationship, to RP peers from other domains that emerge as AS neighbours through MN's current R-neighbourhood. As a result *introduced* RP peers from different domains, operating over the *same* HCoA ranges can effect inter-domain forwarding of the relevant HCoA flows by means of a unicast tunnel.

The peering relationship emerges *dynamically* from the early stages of HAR discovery (HARD). Whether through manual configuration or dynamic intra-domain bootstrap, PoAs within a domain identify their RPs. An *RP peering* between two domain neighbours that have PoAs within the same R-neighbourhood accommodating the MN, is effected during HAR discovery; a Handoff AR update is augmented to include also the identity of its *RP peer* operating over MN's particular HCoA address range. It is intuitive that for a small, manageable number of RPs apportioning satisfactorily the total number of operational HCoAs into managed ranges, PoA neighbours can inform each other about the entire set of RP peers with a *very small* number of messages, effected proactively.

PoAs acting as the designated router (DR) *forward* the identity of the established RP peer for automatic inter-domain bridging for the purposes of inter-domain HandoffCast forwarding. By exploiting cross-domain consistency of the multicast HCoA allocation process described earlier, packets can be forwarded transparently once HandoffCast is initiated, over inter-domain tunnel bridges to the relevant RPs and still get delivered to cross-domain PoA neighbours, while the MN's is executing its IP handoff.

5.4.6 HCoA membership management

Once allocated, the HCoA address is attached onto the IP-Roaming state prepared by the current PoA (AR_c) (proxy-stateless generation), before sent via the respective CtS-Response back to the MN.

From the perspective of HandoffCast management, once the MN receives its IP roaming state, sends an *explicit join* Multicast Listener Discovery (MLD) membership Report [321] to its AR_c for that HCoA address. An explicit join on the part of the MN, is required to ensure that the MN configures the multicast filter on its hardware interface [109], while the receiving AR enables multicast forwarding for that HCoA on the local link.

AR_c requests further that members of its virtual R-neighbourhood join the same HCoA for the purposes of MN's handoff. Members of the R-neighbourhood must enable forwarding of traffic destined towards the MN's multicast HCoA in their downstream interface. This is because *tentatively*, there exists interest in that traffic by at *exactly* one host identified by the soft CoA already allocated.

Under traditional IP multicast semantics, traffic destined to a multicast group (address) is forwarded in the downstream interface of an AR, if there exists interest by *at least one* host. This is the case during HandoffCast forwarding; there exists *exactly* one receiver: the roaming MN.

Despite the MN's residence on a different link, the soft CoA allocated for it proactively at the new AR, makes a statement of tentative but imminent existence for that MN: *the MN is soon to exist on that network link and is interested in the particular group traffic.* In fact, in this light HandoffCast may provide also support for native IP-Multicast communications over mobility enabled IPv6 wireless networks; this is, however, beyond the scope of this thesis.

The request of AR_c onto members of its R-neighbourhood is effected by sending to each AR neighbour an *implicit join* or *I-Join* message; this signal triggers the normal join process by the AR neighbour, according to the multicast listener discovery (MLD) standards [322, 323, 324]. In effect, the MN joins *indirectly*, under HandoffCast management, its multicast HCoA address at PoA members of its R-neighbourhood, which are candidate for its next IPv6 handoff; that is, it expresses interest in its own traffic at candidate points of attachment during its next IPv6 handoff. The devised multicast HCoA allocation algorithm ensures that such interest remains *unique.*

The explicit join of the MN, is effectively *solicited* by the AR_c, through the dispatch of the IP Roaming state. The MN must join the HCoA address, if flow forwarding is to be supported during its next IP handoff. For reasons of signalling economy, the solicitation is piggybacked in the CtS-Response message to the MN by means of a join-bit flag (J).

The purpose of a solicitation is *urgency*; the MN must *immediately* join its HCoA address, once it has received its Roaming state for the purposes of its next IPv6 handoff. The reason of urgency for a roaming MN is multi-fold: (i) IP roaming state is critical for handoffs and thus, must be in place before other state (such as AAA or QoS) can be also proactively established, (ii) the residence period of the MN within an CA is non-deterministic, (iii) cell-bounce (aka ping-pong) effects can occur.

An *implicit join* can implemented in two possible forms: (i) *persistent* and (ii) *periodic.* A persistent I-Join message is included in the CtS-Generate message, sent by AR_c to the AR neighbours during state establishment time. It is essentially a flag and encompasses also MN's HCoA that the AR neighbour must join. This piggybacked signal requires no timer setup on the part of the AR neighbours and thus, no periodic refreshes; instead AR_c explicitly requests the particular AR neighbour to leave MN's HCoA group where necessary through an implicit *I-Leave* message.

Persistent I-Join/Leave handshakes simplify significantly the implementation of the HandoffCast mechanism in contrast to their periodic counterparts. This is because: (i) MLD reports are signalled indirectly for the purposes of HandofCast management, (ii) a persistent MLD membership report scales better for large number of HCoAs. This is because it economises on signalling per HCoA, given that every HAR maintains and signals to its own R-neighbourhood. Under a periodic MLD reporting regime, *every* last hop AR would incur a fixed signalling overhead simply for the purposes of keeping their HandoffCast neighbours alive per HCoA address.

For a persistent I-Join message, we extend the current IGMPv3 specification [322]; AR_c transmits a modified[14] MLD membership report to each of the candidate AR_n neighbours, *on behalf* of the MN as a tentative visitor for each candidate AR_n link.

Each of the PoA neighbours, receiving an I-Join message, enables multicast for-

[14]Only by means of a persistence flag

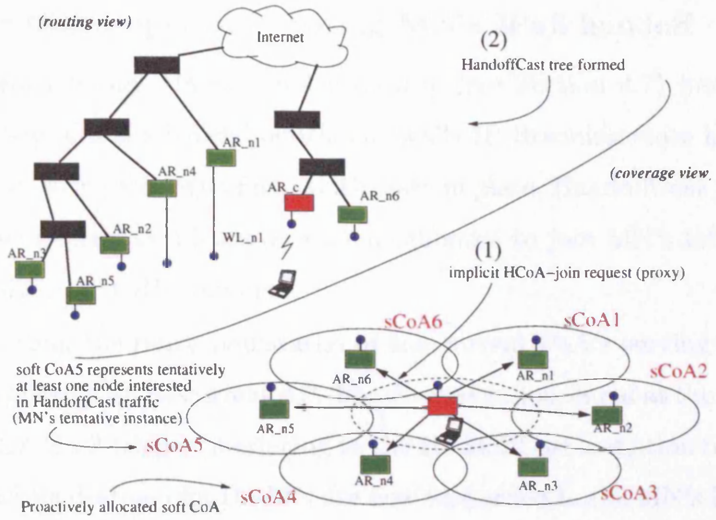


Figure 5.8: HCoA membership of AR_n neighbours is managed implicitly by the current PoA (AR_c)

warding for the HCoA address over that link interface where the soft CoA generated is topologically correct. The PoA neighbour enables a join to that HCoA, while defending for mobility management purposes the soft CoA already allocated for the MN. Figure 5.8 illustrates a simplified spanning tree topology over which an AR neighbour (e.g AR_{n6}) is instructed to manage HCoA group membership through implicit persistent MLD reports from the current PoA (AR_c). It becomes apparent that HCoA membership management is performed through indirect signalling via MN's current PoA.

Flow re-routing under HandoffCast is enforced only for the period of MN's IP handoff. When the MN does not engage in an IP handoff, the HandoffCast function is suspended. Suspension of traffic forwarding over the HCoA group *does not imply teardown* of the delivery tree for MN's HCoA, at AR_c. This is because IP traffic *pursues* its destination host (i.e. the MN), through the corresponding HCoA, *on a per-handoff basis.*

Instead, the multicast delivery tree is managed by having both previous and new PoA signal selected HAR neighbours, the explicit expiry of their membership for that HCoA address. In this manner, HandoffCast ensures minimal multicast tree reconfiguration. It follows that through such approach, the RP effectively *tracks* the movement pattern of the MN by adjusting its HCoA delivery tree to current sCoA receivers.

5.4.7 HandoffCast operation during MN's IPv6 handoff

HandoffCast assumes that Handoff AR discovery (see Section 4.7) has established a complete mapping of MN's R-neighbourhood, while IP Roaming state has been established proactively with its current PoA. With state in place, HandoffCast prescribes that the current PoA invites the identified PoA neighbours to join MN's HCoA indirectly, in advance of MN's next IPv6 handoff.

Upon exceeding the range boundaries of the current PoA's serving coverage area, the MN de-associates with its current AP. The de-association signal at the AP subsystem initiates the `DETACH` L2-trigger interfacing to the HandoffCast initiation function of that PoA. Hence packets destined for the MN are now forwarded to the MN's HCoA address.

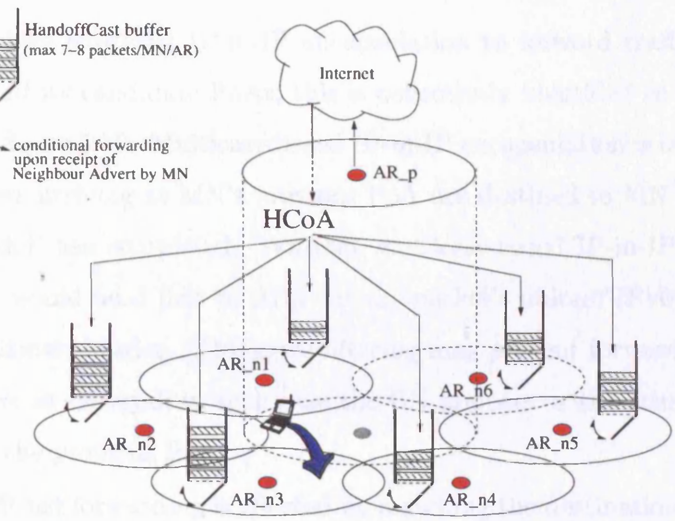


Figure 5.9: HandoffCast buffer configuration at candidate PoA

During this period, the MN is already actively scanning for new PoAs through a link-layer handoff. Link-layer AP discovery maps through the respective ($APID_i, Channel_{APIDi}$) L2-hint the associated AP onto MN's new primary CoA address from the sCoA tuple candidates (see Section 4.8.4). Throughout the period of MN's handoff, packets arriving at a new PoA neighbour are stored temporarily on a circular buffer of about 200ms. The configuration of HandoffCast buffers at the candidate PoAs is shown in figure 5.9.

As soon as the MN completes the association stage of the L2-handoff, the AP subsystem initiates the respective `ATTACH` L2-trigger that: (i) removes the `PROACTIVE` flag from the tentative neighbour cache entry stored at the new PoA during sCoA configuration time (ii) instructs the AP buffer to forward the stored packets over the wireless link and deliver them to the MN on-link This forwarding is enabled through the standard motions of a reachable neighbour cache entry as per the Neighbour Discovery standard [107].

As the MN has effectively completed its IPv6 handoff by means of the **ATTACH** L2-trigger established by the PoA, it can update immediately its bindings by sending a binding update (BU) to its peers, both HA and CN(s). This is followed by an *HCoA-Disable (HCoA-D)* message to the AR_p, through its new primary CoA, to request suspension of HandoffCast forwarding through its HCoA address. The HCoA-D message causes the previous PoA to (i) stop forwarding over the HCoA address (ii) inform neighbourhood-redundant PoA neighbours to leave MN's HCoA address.

HandoffCast forwarding Considerations

HandofCast, employs *multicast* IP-in-IP encapsulation to forward traffic destined for the MN towards *all* its candidate PoAs; this is collectively identified as a *HandoffCast Tunnel*, shown in figure 5.10. Multicast-based IP-in-IP encapsulation is essential for two reasons: (i) packets arriving at MN's previous PoA are destined to MN previous CoA, before MN's handoff has completed. Without multicast-based IP-in-IP encapsulation the previous PoA would need first to strip out the packet's unicast IPv6 header, before attaching the multicast header, (ii) ingress filtering may prevent forwarding if only the destination address is changed; in such case the CN appears as the source, as opposed to the address of the previous PoA.

IPv6 HandoffCast forwarding is effected by matching the destination address in the packet with a *HandoffCast Cache (HC-C)* entry mapping to MN's respective HCoA.

We note that *arrival* of HandoffCast packets at the receiving MN does not depend on any unicast (current or tentative) sCoA allocated for the MN; instead the configuration of the hardware interface (link-layer) multicast filter at the MN, depends only on the last four octets of the HCoA. Thus, all is required between the MN and an AR neighbour is link-layer connectivity together with (link-local) neighbour reachability.

A HandoffCast tunnel requires the use of a HandoffCast tunnel *(HC-T) flag* placed as a destination option in the outer IPv6 header of the encapsulating packet. The AR_c must also mark the `<Next Header field>` within the encapsulating UDP header, with a special type that is called `IP_ENCAP` and denotes an encapsulated packet as the payload of a UDP header.

On receipt of the HandoffCast packet, the MN checks whether the HC-T flag has

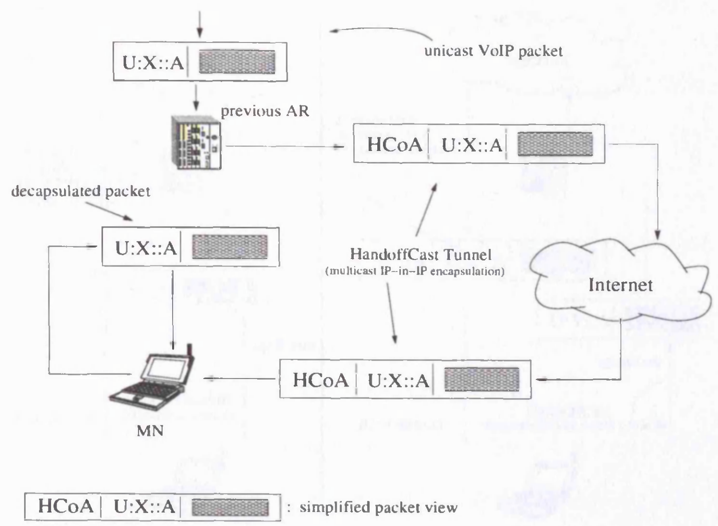


Figure 5.10: HandoffCast tunnel during MN pursuit of downstream traffic under Proactive handoff management

been set in the destination options of the IPv6 header; in addition the MN checks the UDP header, whether the `<Next Header field>` has been set with the `IP_ENCAP` value denoting UDP encapsulation of an IP packet. If this is the value of the next header field in the UDP header then, the decapsulated packet is then re-submitted back to the IP stack.

However, the packet would now have as destination address MN's *previous* CoA configured at AR_p; for this reason,the MN must keep its previous CoA active[15] until its has updated its new bindings with its peers; that is, during the period of its IPv6 handoff the MN maintains an interface configured with *two* CoA addresses: (i) its previous CoA, required for HandoffCast *downstream* communications, (ii) its new primary CoA, required for normal *upstream* communications with its peers. The handoff period is terminated by MN's bindings update to its peers.

We note that such temporal usage of two CoA by the MN does not affect routing towards the candidate PoAs of the MN in any way. For the period of the handoff, the MN simply de-multiplexes bidirectional communication with its peers, through two identifiers: the previous CoA for downstream packet reception, and the new primary CoA for upstream communication. The notion of stream-demultiplexing under HandoffCast is illustrated in figure 5.11.

[15]A single NIC interface can be configured with more than one IPv6 addresses.

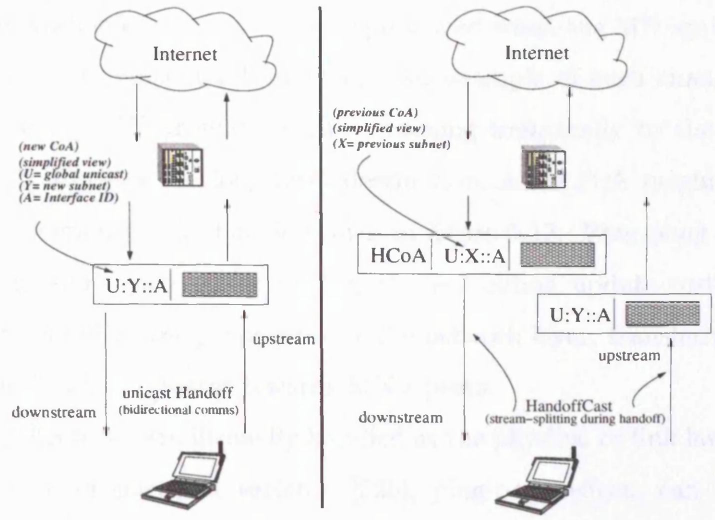


Figure 5.11: Stream-de-multiplexing under HandoffCast. The MN exploits temporally its previous CoA to de-tunnel HandoffCast packets within its network stack. Downstream routing remains unaffected.

5.4.8 Managing network-layer ping-pong effects

Typically, the MN updates its peers with its bindings, whenever it moves to a new PoA. Excessive updates may, however, occur when the MN's handoff rate increases as a result of *ping-pong* effects [325]. A ping-pong effect occurs inherently[16] in direct sequence (DS) CDMA systems as a result of multiple-access interference (MAI) [326].

A wireless MN receiver employs typically interference cancellation, to discriminate between the receiving and interfering signals of multiple APs. To achieve this, the detector employs normally iterative techniques[17] for (signal) matrix inversion [327]. Well-behaved reception achieves usually signal detection within a few[18] iterations.

A ping-pong effect occurs as a result of a tendency of the iterative method, for the sampled bit-error rate (BER) of the detected signals, to alternate between two states as a function of the iteration index: the BER envelope for the odd appears to be better that the envelope of the even iteration per signal [304]. When the signal is affected by propagation effects (e.g. path loss, multi-path fading), the detector decision oscillates between the 'best' BER experienced by receiving and interfering (but competing) signals and thus triggers a link-layer handoff.

Propagation effects are in turn sensitive to MN's movement trajectory and propagation obstructions. Ping-pong effects emerge typically at the boundaries of coverage

[16]based on the impossibility of maintaining spreading code orthogonality
[17]e.g. Jacobi iteration
[18]as many as the size of the signal matrix

areas. However, such effects may also be experienced when the MN appears to be well within the range of a particular PoA [306]. An example of such change of direction is the turning of the MN around corners returning temporally to the previous PoA before continuing towards the long-term destination, while PoA neighbours maintain some minimum overlap [328]. This is shown in figure 5.12. Ping-pong induced handoffs in cellular systems account for 15-22% of the location update cost [305, 236]. It is intutive that a similar cost propagates at the network layer, translating for mobility management, to binding updates towards MN's peers.

Ping-pong effects are traditionally handled at the physical or link layer. Depending on the measure of propagation variance [329], ping-pong effects can be reduced by introducing a measure of hysteresis at the physical/link layer whereby a conservative power (link) margin [330] of the wireless link is allowed to fall below a hysteresis margin of a few dB. Typical power margins for a wireless link have values of around 3-7 dB.

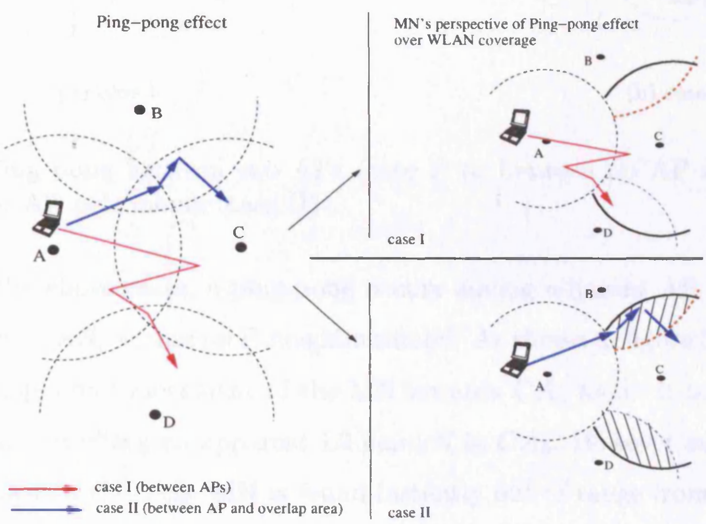
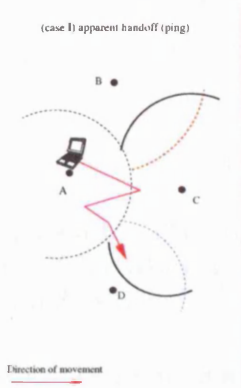


Figure 5.12: Ping-pong effect experienced by the MN amongst neighbouring APs.

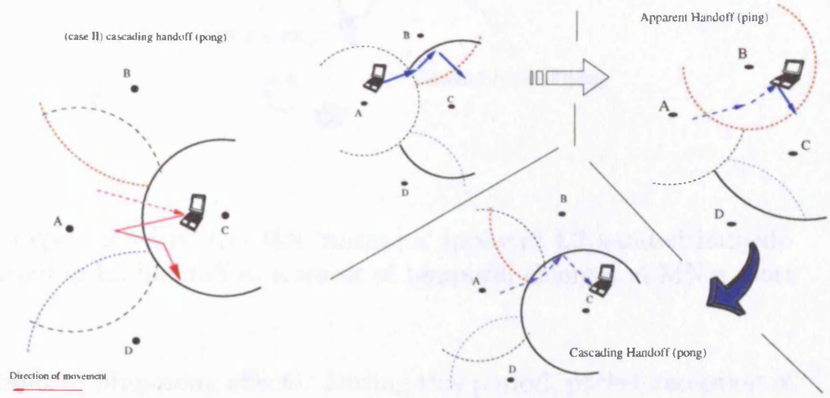
Through a hysteresis margin at the MN, the wireless link caters effectively for a *grace* period (which is dependent on MN's velocity) within which the MN may either return back towards its current cell centre or move over the neighbouring CA. For small propagation variance $\sigma \leq 7$, hysteresis schemes perform satisfactorily with a typical hysteresis coefficient of $h = 3dB$ [331].

From the perspective of the network layer and subsequently IPv6 handoff management, a ping-pong effect implies cascading IPv6 handoff between two AR neighbours in two distinct cases: (i) one where the MN oscillates fast between the previous AR_p

and the new PoA AR_n, (ii) one where MN's mobility pattern incurs a cascading IPv6 handoff between the previous AR_p and two different neighbouring new ARs AR_n and AR_{n+1}. In both of these cases, shown in figure 5.13, cascading IPv6 handoffs between two PoAs translate to an increase of MN's binding updates to its peers. It is essential for the purposes of IPv6 mobility management to reduce unwanted binding updates until temporal PoA handoff oscillations have *settled.*



(a) case I



(b) case II

Figure 5.13: Ping pong between two APs (case I) or between an AP and an overlap area of two new AP neighbours (case II)

For both the above cases, a ping-pong occurs among adjacent AR neighbours directly reachable by AR_p within its R-neighbourhood. As shown in figure 5.14 in a case-II ping-pong, the apparent movement of the MN towards CA_b forces it to experience an $SNR_B > SNR_C$, resulting an apparent L2-handoff to CA_b. However as the MN turns momentarily towards CA_c, the MN is found instantly out of range from CA_b resulting a cascading L2-handoff to CA_c

For mobility management mechanisms that do not explicitly support network layer ping-pong effects in either of the ping-pong cases, the above would result two consecutive binding updates towards all of MN's peers and thus double temporally the location update cost. Under HandoffCast, the MN continues to receive its traffic over its HCoA address until it attaches to some PoA neighbour, plus a small time period T_e[19]; this period is identified as *network-level (L3) hysteresis* for the purposes of IP handoff completion.

An L3-hysteresis aims to reduce the temporal number of BU signals sent by the

[19]defaults to 1000-1500ms

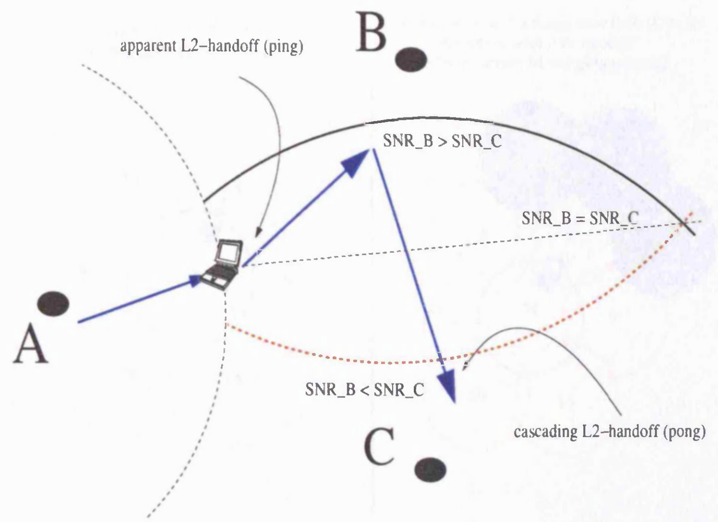


Figure 5.14: The SNR experienced by the MN incurs an apparent L2-handoff immediately followed by a cascading L2-handoff as a result of temporal changes in MN's short term direction.

MN to its peers as a result of ping-pong effects. During this period, packet reception of MN's traffic is sustained through its HCoA address.

The MN maintains also a measure of its handoff rate compared against a threshold of 1 handoff/sec after each handoff; if the rate exceeds this threshold, then the MN increases exponentially its L3-hysteresis period up to an upper bound[20] of 5000ms. A similar technique is adopted in modern cellular (UTRA/UMTS) systems for cells that belong in the *Neighbour Set* as likely candidates for the *Active Set* during a soft handoff [332].

Exceeding the handoff rate threshold is guaranteed to be caused only by temporal changes in MN's mobility pattern, locally within its M-neighbourhood, as shown in figure 5.15; this because, the MN can only exceed such handoff rate either:

- due to temporal changes in MN's direction with respect to the cell boundaries of APs *within* its current M-neighbourhood.
- due to a magnitude of velocity that on average is greater or equal to transmission range of the AP at the *new* PoA. In this case the MN must interact with APs *outside* its current M-neighbourhood.

To assess the possibility of occurrence for the second case, we evaluate the magnitude of MN's velocity under the transmission range specification typically met in

[20] `Max_Random_HCoA_Rx Stop_Delay` in specification [129]

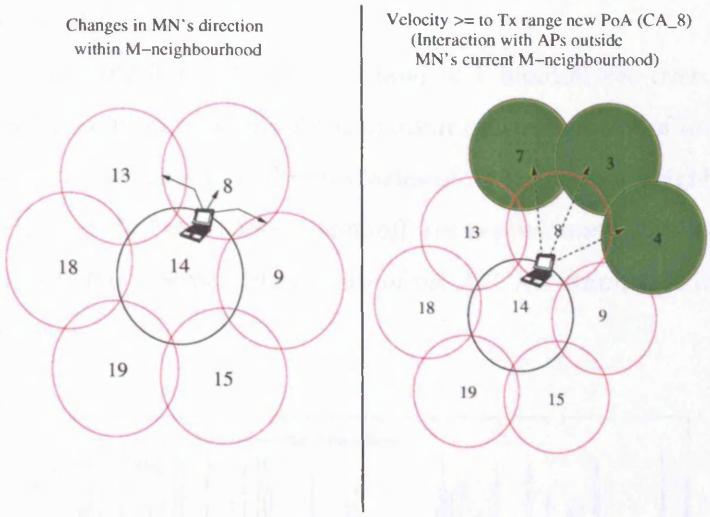


Figure 5.15: The improbable case of a handoff rate ≥ 1 handoff/sec due to velocity.

IEEE802.11 vendor implementations. This may be found in Table E.1, Section E.2. If one assumes that each of the transmission ranges of Table E.1 define the boundaries of a WLAN coverage area, then Table 5.1 shows the hypothetical velocities required by the MN to ensure a handoff rate equal to 1 handoff/sec.

propagation \ B/w (Mbps)	11	5.5	2	1
	km/h			
Obstructed environment ($V_{indoors}$)	90	126	144	180
Semi-obstructed ($V_{in/outdoors}$)	180	252	324	414
Free Space ($V_{outdoors}$)	576	972	1440	1980

Table 5.1: Velocities required by the MN to ensure a handoff rate of 1 handoff/sec at the coverage boundary of each signalling rate. Such velocities are clearly unattainable and thus, render the 1 h/sec threshold due to MN speed, strongly improbable; for instance, for an indoor propagation environment the MN would require a velocity of 144 km/h to effect a handoff rate of 1 h/sec.

For all signalling rates, each defining its effective cell boundary[21], the MN is required to maintain vehicular speeds that are *strongly improbable* to attain over most terrain environments.

For instance, for obstructed indoors environments, the MN can only experience a handoff rate of 1 handoff/sec if it sustains an average velocity of 90-180 km/h, which is clearly impossible to attain. In a similar fashion semi-obstructed (city) or free space propagation environments (country/suburban) require that the MN maintains average velocities above 180 km/h, clearly an impossible case given current vehicular speed

[21] at the particular sensitivity threshold.

limits[22] and terrain obstructions.

Hence, exceeding the handoff rate threshold of 1 handoff/sec over administrative domain that span horizontally is guaranteed to occur only as a result of temporal changes to MN's direction with respect to cell boundaries of its current M-neighbourhood. The above shows that, *a handoff rate above 1 handoff/sec is predominantly owed to ping-pong effects, typically near the coverage boundaries of the AP*. An example of its measure may be seen in figure 5.16.

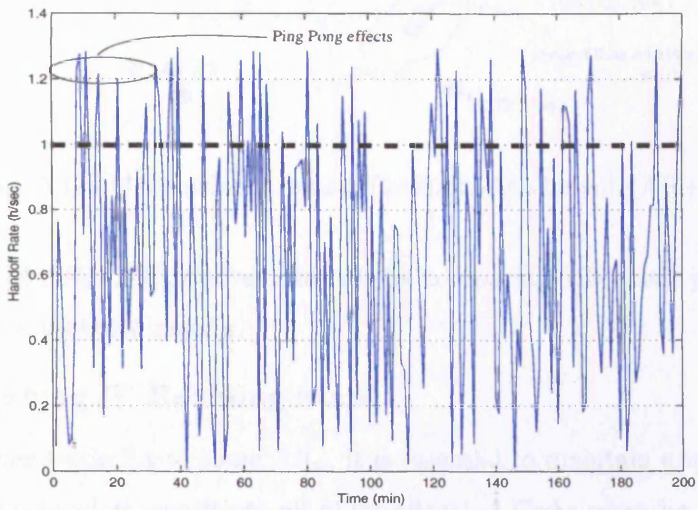


Figure 5.16: Ping-Pong effects emerging from measure of MN's handoff rate

Other causes of a handoff rate above 1 h/sec is the case of vertical handoffs [149]. It is reminded that an M-neighbourhood operating under HandoffCast forwarding, is not limited by neither horizontal nor vertical service provisioning. Hence, the MN is able to effect a vertical handoff, subject to a minimum cell residence period and thus, yield a higher handoff rate.

With respect to handoff completion, when the L3-hysteresis period elapses, while MN's handoff rate remains below 1 handoff/sec, as shown in figure 5.17, the MN sends a standard BU to its peers (CNs/HA) to inform about the new primary CoA. At the same time the MN sends a *HCoA-Disable (HCoA-D)* message to AR_p, through its new primary CoA, to request suspension of traffic forwarding through the HCoA.

The HCoA-D message instructs further AR_p to manage group membership of the HCoA address; it includes 'pruning' of PoA neighbours that do not belong in the MNV-

[22]It is important to note that for high speed vehicles, such as aircrafts, different types of wireless technologies and thus transmission ranges would be applicable

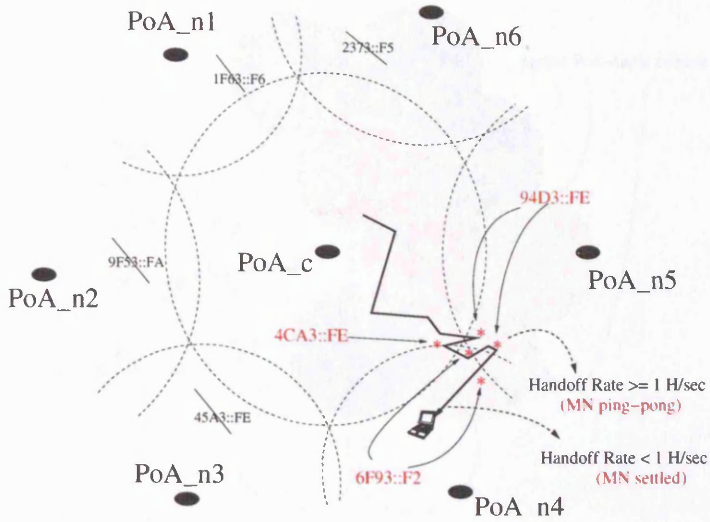


Figure 5.17: MN settles at a new Point of Attachment (AR+AP)

RNV mapping of AR_n. AR_p derives such set by comparing MN's new primary CoA[23], with its tentative mobility matrix.

5.4.9 Refreshing IP Roaming State

Once the MN has settled onto some AR_n, it is essential to maintain a *valid* PoA tuple, for its next IPv6 handoff transition; all of its allocated CoAs must be valid within its *new* current R-neighbourhood.

To that effect, the MN requests from the new PoA (AR_n) an IP Roaming state update by means of a standard *CtS-Request* message with the `update` flag set; Such request initiates on the part of the new AR, a PoA tuple *refresh* for that MN, encompassing a IP-Roaming state *delta* between the R-neighbourhood of MN's AR_p and AR_n, identifying the *new* AR neighbours.

ARs in the R-neighbourhood of AR_p maintaining group membership in MN's HCoA group address leave the multicast tree, by means of an explicit request by the previous PoA. The new AR_c is aware of the LLA of the MN, since it created a soft CoA for that MN during its movement at AR_p. Hence, the new AR_c simply confirms the CtS-Request based on trust, built during IP-Roaming state generation delegated in MN's previous handoff by the previous PoA (AR_p).

A refresh of IP-Roaming state in the M-neighbourhood is depicted in Figure 5.18; the schematic assumes an M-neighbourhood of 6 PoAs; however, the principle can be

[23]source address of the HCoA-D message

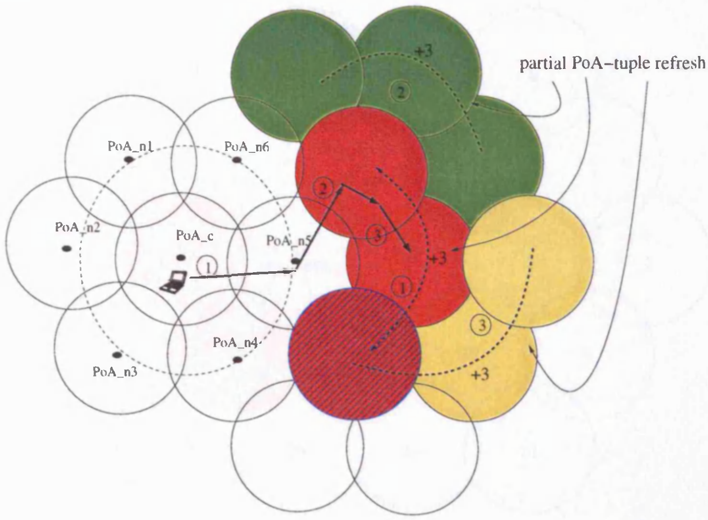


Figure 5.18: sCoA-tuple reuse within the MN's PoA tuple during continuous movement

generalised for any M-neighbourhood size. During the first IP Roaming state allocation the MN receives a PoA tuple with *six* new sCoA-tuple constituents. After its first IPv6 handoff, MN's PoA tuple need only be updated with another *three* sCoA-tuple constituents. For a random PoA neighbourhood size, an IP Roaming state update requires in general, about *half* the number of soft CoAs allocated during MN's first IPv6 handoff.

In the next subsection we present the algorithm in which resolution of redundant and new AR neighbours is performed. It encompasses further the PoA tuple update, that allows the MN to maintain valid IP-Roaming state during its next IPv6 handoff.

Resolution of the PoA-tuple delta

The *CtS-Request* update message sent by the MN includes also the address of the previous PoA (AR_p) in a manner similar[24] to MN hints under HARD, presented in Section E.3.2. On receipt of this message, the new PoA (AR_n) determines the AR neighbours *common* to both AR_n and AR_p, by checking the address of AR_p against the contents of its tentative mobility matrix (see Section E.3.3), shown in Figures 5.19 and 5.20. Once the common PoAs have been determined, AR_n initiates generation of the IP Roaming state *update delta* for the new PoA neighbours. Such signal directs implicitly the *new* AR neighbours to join MN's existing HCoA address.

Furthermore, the HCoA-Disable signal sent by the MN, directs AP_p to prune from

[24]The signal may be optimised by combining M/R neighbourhood discovery and context state establishment (CtS-Request) into a single control signal

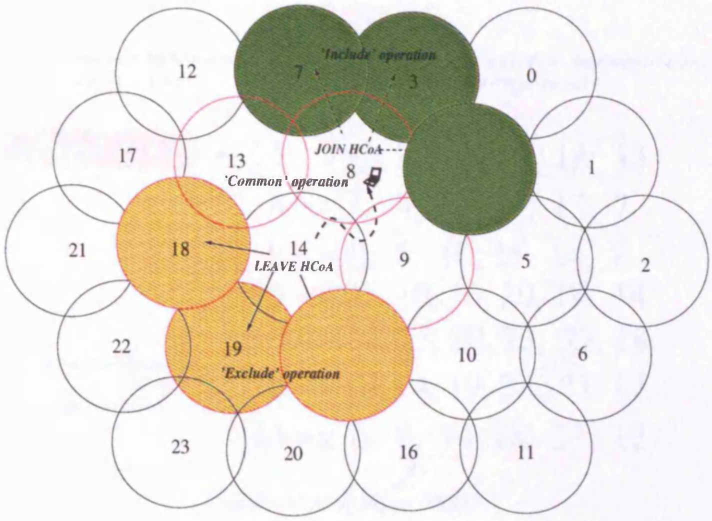


Figure 5.19: Sustaining accurate mapping of AR neighbours on MN's HCoA group routing identifier

MN's HCoA address, any *redundant* PoAs; these are PoAs that are not found in MN's *current* R-neighbourhood.

Pruning of redundant PoAs is achieved by having the previous PoA send an *implicit leave (I-Leave)* to these PoA neighbours; the PoA neighbours common to the R-neighbourhood of both AR_p and AR_n remain onto the HandoffCast tree. The I-Leave signal, flags further the release of redundant IP-Roaming state, at redundant AR neighbours, namely addressing and reachability state pertaining to the sCoA tuple configured previously for the MN.

Derivation of the PoA-tuple delta is performed by means of 3 simple set operations: *common, include, exclude.* The PoA tuple update operations are defined as follows:

$$Common(CtS_p, CtS_n) = \bigcap(CtS_p, CtS_n) \quad (5.3)$$

$$Include(CtS_p, CtS_n) = CtS_n - Common(CtS_p, CtS_n) \quad (5.4)$$

$$Exclude(CtS_p, CtS_n) = CtS_p - Common(CtS_p, CtS_n) \quad (5.5)$$

where $rnv_{p,n}$ are the RNV vectors of AR_p and AR_n maintained in their RNV Cache. Figure 5.20 illustrates the tentative mobility matrix maintained in the RNV-Cache of an AR; according to the above operations, the resulting vectors V_{common}, $V_{include}$ and $V_{exclude}$ take the following sample values according to Figure 5.19:

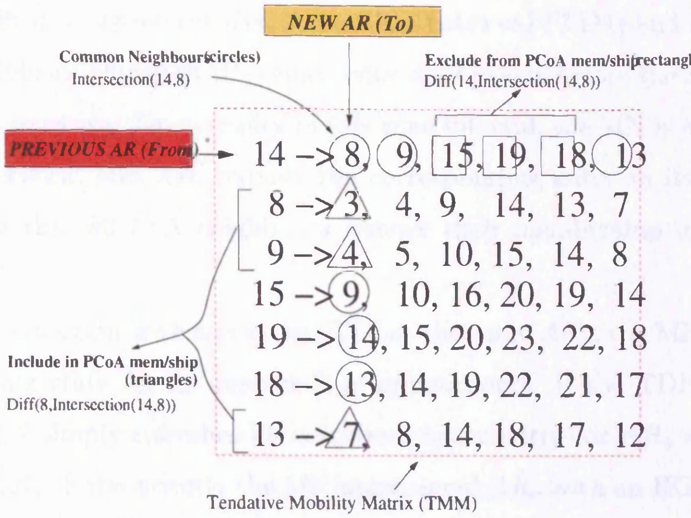


Figure 5.20: Resolving redundant AR participants through the tentative mobility matrix. The previous PoA determines the set of PoAs that are *excluded* from MN's current R-neighbourhood according to Eqn. (5.5). The new PoA identifies the set of PoAs that are *included* in MN's current R-neighbourhood according to Eqn. (5.4).

$$CtS_p = CtS_{14} = \{14, 8, 9, 15, 18, 19, 13\}$$
$$CtS_n = CtS_8 = \{8, 3, 4, 9, 14, 13, 7\}$$
$$CtS_{Common(CtS_{14}, CtS_8)} = \{8, 14, 9, 13\}$$
$$CtS_{Include((CtS_{14}, CtS_8)} = \{3, 4, 7\}$$
$$CtS_{Exclude((CtS_{14}, CtS_8)} = \{15, 18, 19\}$$

Forced disruption of IP connectivity

It is possible that due to limitations in wireless coverage, the movement pattern of a MN may enforce a disruption of IP connectivity in its current mobility neighbourhood; typical example is movement through road tunnels, or underground train stations. In these cases, the MN appears from the perspective of the network potentially as (i) performing a handoff or (ii) unreachable. The second case may also give rise to the scenario where the MN resumes IP connectivity at some AR out of the current PoA's M-neighbourhood.

In the event of a forced disruption of IP connectivity, the current PoA initiates HandoffCast assuming the MN has performed an IPv6 handoff. Past the L3-hysteresis period, the current PoA, delays flow forwarding for an additional time T_{tdi}; this time period is identified as **transient disconnection interval (TDI)** and represents the interval permissible for transient IP connectivity disruptions before the MN is considered out of network coverage. Upon expiry of this time interval, the MN is assumed unreachable by the network, and AR_c expires the corresponding entry in its CS-Cache; AR_c ensures further that all PoA neighbours remove their membership from MN's HCoA address.

Upon re-connection with an expired TDI at the same AR_c, the MN must re-acquire new IP Roaming state for its current R-neighbourhood. If the TDI interval has not expired, the MN simply refreshes its neighbour cache entry for AR_c as REACHABLE according to [107]. Subsequently the MN must signal AR_c with an HCoA-D to suspend flow forwarding.

In the event that the MN re-connects with an AR different than AR_c, the MN must discard its existing IP Roaming state and provide its link layer address to the new AR_n for the acquisition of fresh IP Roaming information.

5.4.10 Delay Seamlessness and Security Considerations

Most IPv6 mobility management standards currently proposed by the IETF, mandate MN authentication mechanisms common to the ones of the Mobile IPv6 standard. This is the case for both Fast Handoffs [33] and Hierarchical MIPv6 [301] specification or their derivatives [300].

One of the most important issues in terms security under IPv6 Mobility management is authentication of Binding Updates or control signalling in general, sent to peers by the MN. To this end, the Mobile IPv6 standard (and through it, most popular mobility management mechanisms) identifies a secure mechanism of binding updates identified as *Return Routability (RR)* described briefly in Section D.1.6.

The RR mechanism enforces authenticity of the MN as the true originator of a Binding Update sent to its CN peers, by means of a challenge-response test, initiated by the MN and conducted by the CN through two paths: (i) via the HA, (ii) through the route-optimised path between CN-MN (i.e. direct to MN's new CoA). Such means of authentication, however, introduces significant latency in cases of increased RTT, as discussed in Section 3.7.1. The RR function incurs a BU delay component that is at least two times longer than that of an unauthenticated BU. This is the case for every handoff the MN performs between PoAs and applies for all existing mobility management extensions including as Fast handoffs and HMIP.

In an effort to reduce delay introduced by such security mechanisms, Proactive IPv6 mobility management follows a different approach to the RR mechanism; authentication of BU control messages is enforced by means of public-private key non-repudiation [333] applied onto an extended form of Binding Updates, identified in Proactive IPv6 mobility as *Signed-Binding Update (S-BU).*

The mechanism removes the need for multiple cryptographic generation functions on the MN by placing the requirement for a single light-weight[25] certificate authority that may be co-located on the HA or on-link with the HA host. To simplify the description of the mechanism we assume that such certificate authority is co-located on the HA host; the mechanism assumes that the HA host is secure.

In advance of its IPv6 handoff, the MN sends proactively a *Non-repudiation Qualifier (NRQ)* message to the *communicating*[26] CN. The message includes a *digital signature* [333] (generated by the MN) and the address of the certificate authority (the HA) where MN's public key may be obtained.

Once the CN receives an NRQ message, it extracts the NR-signature and contacts the HA through a *PK-Request*, requesting the public key for the source address of the node. The CN then receives MN's public key, by means of a digital certificate; it then performs a *test-MN-verification* on the extracted signature by using MN's public key. If the test is successful then the public key received by the HA (and compared with the one sent by MN), maps to one and only one address: the home address of the MN.

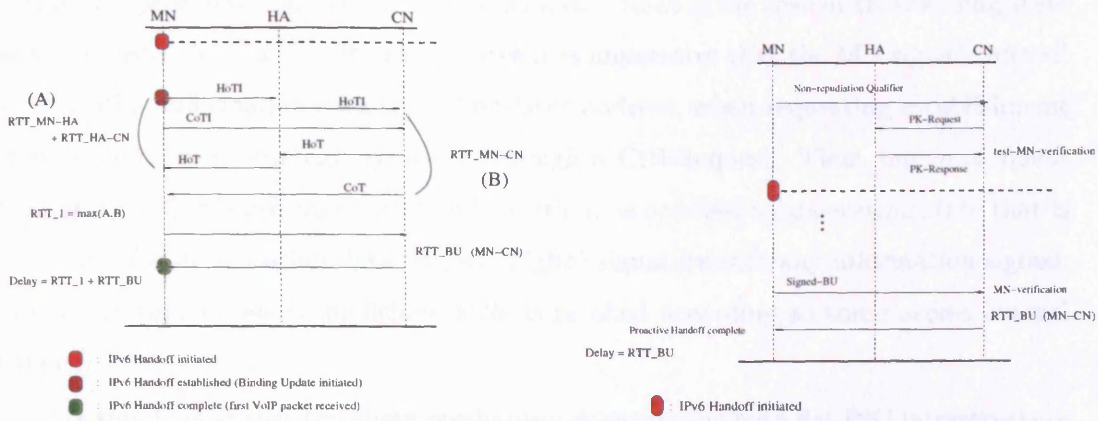
After the completion of its IPv6 handoff the MN sends a Signed-BU to the CN. The MN signs the BU and includes the signature into the NR option. The CN receives the BU message directly by the MN, extracts its NR signature and conducts an *MN-verification.*

It can be seen that the CN receives MN's authentication credentials with *no* delay incurred *during* the period of MN's binding updates. Hence, the non-repudiation mechanism of signed-BUs introduces no additional delay during the critical period of an IPv6 handoff. On the contrary, Mobile IPv6 doubles the RTT delay of a BU signal if authenticated under the RR mechanism. Figure 5.21(a) illustrates the amount of delay incurred during RR-authenticated BUs under Mobile IPv6, while figure 5.21(b) shows the delay incurred by means of the signed-BU mechanism.

The signature is first prepared by generating a hash of MN's IPv6 address; such

[25]Such a certificate authority can easily accommodate a manageable number (in the order of 1000-10000) of MN certificates that can serve one or more links of a single M-neighbourhood

[26]Speculative NRQs are discarded. The MN has to establish first packet communications with the CN before it can send an NRQ message



(a) RR-auth BUs (Mobile IPv6 or derivatives) (b) Signed-BUs (Proactive Mobile IPv6)

Figure 5.21: Comparative Delay incurred by the RR-authenticated BUs under Mobile IPv6 and signed-BUs under Proactive Mobile IPv6.

hash is identified as *BU-digest*. The BU-digest is then encrypted with MN's *private key*; the encryption output forms MN's signature.

A CN verifies[27] the signature extracted from the NR option on the received packet, by means of decrypting the NR signature with MN's public key, obtained from the HA (certificate authority). The resulting message digest is compared against the BU-digest yielding from the hash of MN's current address. If the two digests agree then the originator of the packet is *assured* to be the MN.

With respect to non-repudiation arising from the allocation of HCoA address Section 5.4.4 proposes the generation of a handoff-hash that may be embedded onto the generated HCoA and associated with the LLA address of the MN, while at its home network.

Generalising signed-messages for critical control signalling

The above mechanisms of signed-BU may be generalised for any form of critical control signalling. For instance, discovery of M/R Neighbourhood mappings discussed in Section E.3.2, exploits MN hints about the IPv6 address of their AR_p. It is possible that MNs could provide false hints about the address of their previous AR maliciously. In such case it may prove beneficial to require the digital signing of such information from the MN, so that the receiving AR can identify accurately nodes that behave maliciously and revoke access privileges accordingly.

In a similar manner, the MN may be required to sign *any* initialiser data for the

[27] whether as a test at the home network or as a result of a BU message by the MN

purposes of proactive context state establishment. Such is the case in IP Roaming state establishment. From a security perspective it is imperative that the MN signs 'claimed' identification information such as its link-layer address, when requesting establishment of state under some context, typically through a CtS-Request. Thus, any subsequent state generated is based on accurate information, as opposed to generating state that is based on false initialisation data. A false digital signature over any information signed, can ensure that access to malicious MNs is revoked according to some access control policy [].

We should note that the above mechanism does not assume a flat PKI infrastructure in the order of millions or hundred of thousands MN certificates. On the contrary, the size of the PKI infrastructure required to manage digital certificates under Proactive IPv6 mobility management can be successfully segmented into a hierarchy of light-weight certificate authorities tracking the number of (or potentially collocated with) Home Agents (HAs) allocated within an administrative wireless network domain.

5.5 Performance Evaluation

In this section, we evaluate the performance of HandoffCast as a complementary function of proactive mobility management in support of delay seamlessness. This evaluation assumes IEEE802.11 as the underlying wireless technology, given the increased measure of delay incurred during a link-layer (L2) handoff.

To this end, Annex F.5 looks first into the performance of the measure of persistent L2-handoff delay, as a result of the cross-layer optimisation of 802.11 scan latency proposed in Annex F.4.

5.6 HandoffCast Performance - Methodology

Having shown that proactive guiding of the AP scanning process, can yield sufficient delay reductions, to allow an 802.11 WLAN L2-handoff to complete in $< 100ms$, we focus our evaluation, by means of NS-2 [250] simulations, on the performance of HandoffCast flow forwarding. In particular, we are interested to know whether HandoffCast can support successfully delay seamlessness by complementing Proactive handoff management over handoff delays beyond the control of the network layer. Such support to be efficient must incur a measure of persistent handoff delay below 200ms.

The above hypothesis rests on the performance of HandoffCast evaluated on three accounts:

- L2-handoff delay performance aided with the aforementioned devised optimisation over 802.11 networks, with a target of $< 100ms$.

- a targeted forwarding delay of $< 100ms$ between the new and old PoA during HandoffCast forwarding.

- a total end-to-end one-way delay between the CN and the MN at the new PoA $< 200ms$, during MN's IPv6 handoff.

Furthermore, we assess the cost of HandoffCast in terms of signalling overheads, for the purposes of supporting delay seamlessness. It should be reminded that, delay seamlessness below the network layer may be afforded only at the cost of additional signalling overhead from a protocol optimisation.

With this in mind, we assess what is the measure of additional signalling required to establish and maintain HandoffCast forwarding for the duration of an IPv6 handoff, on a per MN basis. Such a measure will enable an assessment on the trade-off between delay seamlessness and signalling overhead. It provides further, a control measure towards future work on subsequent optimisations of predictive control on the effective size of MN's underlying R-neighbourhood, that tracks the measure of signalling cost incurred under HandoffCast.

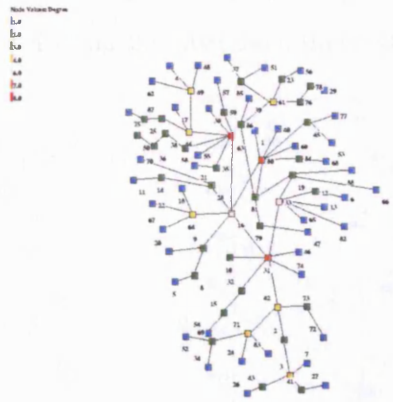
To this end, the simulated wireless network is assumed to operate under average error-free, load conditions. Traffic loading is simulated at 50% of the wireless link capacity (1Mbps), comprizing of a traffic mix[28] identical to the one introduced for experimental measurements in Chapter 3 (see section 3.6.1).

The error-free simplification does not detract from the validity of the proposed mechanism, since all mobility management signals are acknowledged.

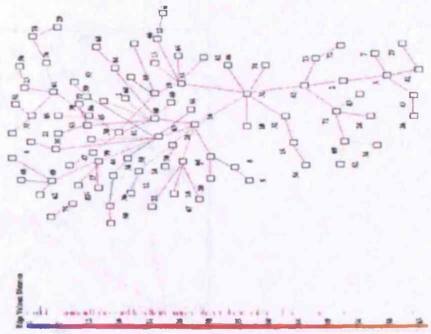
The simulation model employed adopts essentially the methodology of Section 4.9.1, modified by specific HandoffCast extensions. In particular, the L2-handoff delay model is adapted to reflect the aforementioned optimization, with an effective `maxChannelTime` of 8.2ms. The MN implementation is further augmented to support *proactively guided* channel probing, over the set of channels, operating solely *within* MN's current PoA neighbourhood.

With respect to HandoffCast, a non-leaf AR is configured as the RP, fixed on a *random* non-leaf AR node within the topology for a different PoA density. Four PoA

[28]A total set of 4 unidirectional flows: 2 TCP flows at 15 Kb/sec and 1 UDP flow at 10 Kb/sec all destined to the same stationary wireless node in each WLAN cell. The measured VoIP flow consumed 1.62 Kb/sec for a GSM (CBR) encoding at 13.3 kbps.



(a) Node-degree for 45-PoA topology



(b) Weighting over link distance

Figure 5.22: Node degree and associated link distance for generated PoA topology participating in HandoffCast simulations

densities were evaluated, namely 45, 70, 140 and 200 leaf PoAs within the topology. During generation of such set of topologies, we found that on average a set of x leaf PoAs requires the existence of nearly $2x$ (leaf and non-leaf) AR nodes. The random non-leaf AR node selected as the RP for each PoA density was kept fixed for all 20 iterations generation in each PoA density simulation scenario.

Four different node-degrees[29] were recorded during RP placement over an AR node, for the respective four PoA densities experimented; (i) for PoA=45 the RP has a node-degree=8, (ii) for PoA=70 the RP has a node-degree=6, (iii) for PoA=140 the RP has a node-degree=2 and (iv) for PoA=200 the RP has a node degree=3. Figure 5.22(a) illustrates the node-degree of a 45-PoA topology (RP placed on node 63). The underlying multicast forwarding protocol is configured and extended over the bi-directional PIM-SM (sparse mode) NS-2 implementation.

Round trip times are synthetically pre-computed by assigning a one-way delay over the shortest path, between the assigned (fixed) CN and each of the candidate PoAs (leaf ARs) within the entire topology. The measure of one-way end-to-end (e2e) delay is generated from a shifted gamma distribution, according to Section 4.9.1, Section 2.3.3 and Section 2.3.3.

Additionally, nodes with a higher node degree are assigned a *delay weighting*; this reflects the probability of non-leaf ARs with a higher node degree experiencing a higher

[29]Node-degree is defined as the number of connections to other nodes in an undirected graph. For a directed graph it becomes fan-in/out depending on the direction.

traffic load and thus, due to queue build-up, experiencing a higher forwarding delay over the particular link. Such weighting is also influenced by propagation delay of the connecting link distance, illustrated in figure 5.22(b).

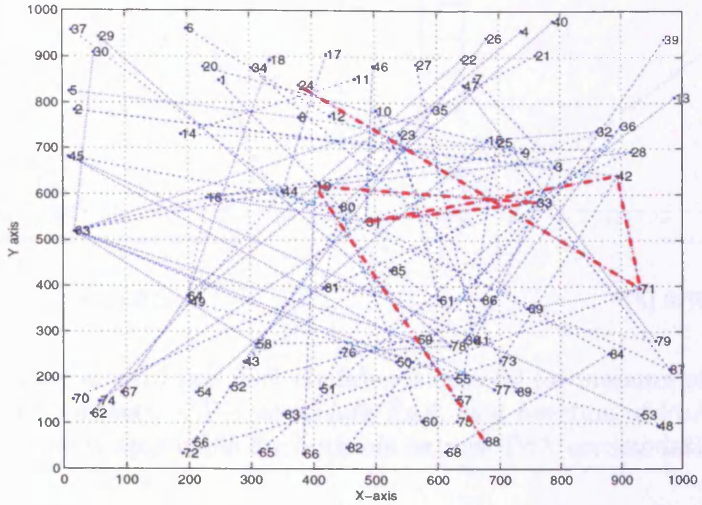


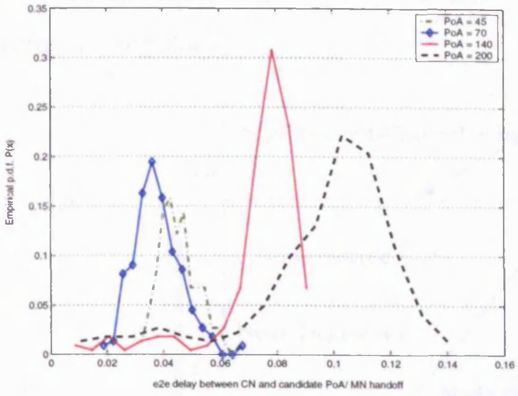
Figure 5.23: Calculating the Shortest Path between the a source and a destination node, using Dijkstra's greedy SP algorithm.

For the measure of one-way delay, the above configuration produces an e2e delay measure, which is differentiated by the size of the PoA topology, as shown in figure 5.24. For larger PoA densities the number of non-leaf ARs within the topology, with a node degree > 1, increases; this results into an increase of high-node degree ARs in the *CN→ old/new PoA* path, which in turn may increase the one-way delay contribution per link traversed due to the above delay weighting.

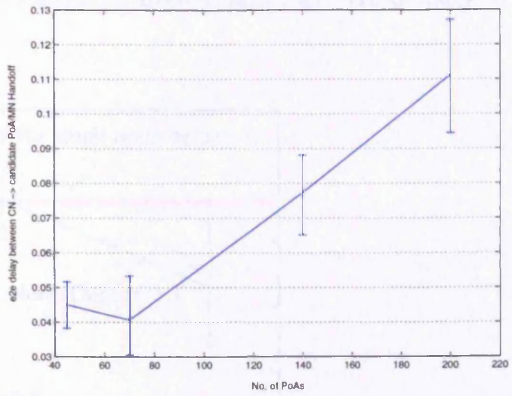
Such link-delay conditioning is employed for the purpose of evaluating HandoffCast forwarding performance under an increasing measure of e2e delay between the CN and candidate leaf-PoAs. The above setting gives an overall *CN→ candidate PoA* e2e delay, independent of the PoA topology, with the following percentiles: $Q_{25} = 35.4ms$, $Q_{50} = 40.7ms$, $Q_{75} = 49.1ms$, $Q_{95} = 83.8ms$ and $Q_{99.9} = 138.3ms$.

Delay over HandoffCast paths is monitored by measuring the one-way delay of packets forwarded *upstream* over the (old PoA→ RP) path and *downstream* over the (RP→ new PoA) path; to achieve this, we compute through Dijkstra's greedy Shortest path algorithm [269, 334], the shortest path between the nodes: (i) old PoA, (ii) RP, (iii) new PoA.

In addition, to assess the measure of total e2e delay as a result of HandoffCast forwarding between the previous and new PoA, we compute further the shortest path



(a) P.d.f P(x)



(b) 95% CI

Figure 5.24: Empirical p.d.f and 95% confidence interval for measure of one-way delay for the *direct* path between *CN→ candidate PoA*, as a function of PoA density. Such measure of e2e delay is applicable for both *old* or *new* PoA accomodating an MN.

between the CN and the old PoA, where the MN is residing. A visualisation of such computation during the simulation is illustrated by figure 5.23.

We note that measurements are traced past the transient period of HARD state convergence, where HandoffCast appears to perform at steady state. To this end, results focus solely on proactive handoffs assisted through HandoffCast flow forwarding.

5.6.1 Simulation Results

Persistent handoff delay and forwarding delay component

Figure 5.25 presents the average measure of persistent handoff delay effected solely on the HandoffCast path between the previous and new PoA, as a function of PoA density, at a 95% confidence interval.

For AR topologies $\geq$ 140 leaf PoAs, we note a steep increase (about 79ms) in the measure of persistent handoff delay from about 80ms to 158ms. Such increase in the measure of this type of handoff delay is due exclusively to the measure of the underlying one-way delay component between the previous and the new PoA, as opposed to increased L2-handoff delay measure. This behaviour can be seen also in more detail in Annex F.7.

Looking at the forwarding delay component of figure 5.26, when packets are path re-routed over HandoffCast towards MN's new PoA, we find that the increase is critically dependent on the choice of AR node placement of the RP. Such rise in forwarding delay, emerges specifically from the *differentiation on node-degree of the router where the RP* *is configured*, rather than a hop increase due to explicit higher non-leaf AR density within the topology.

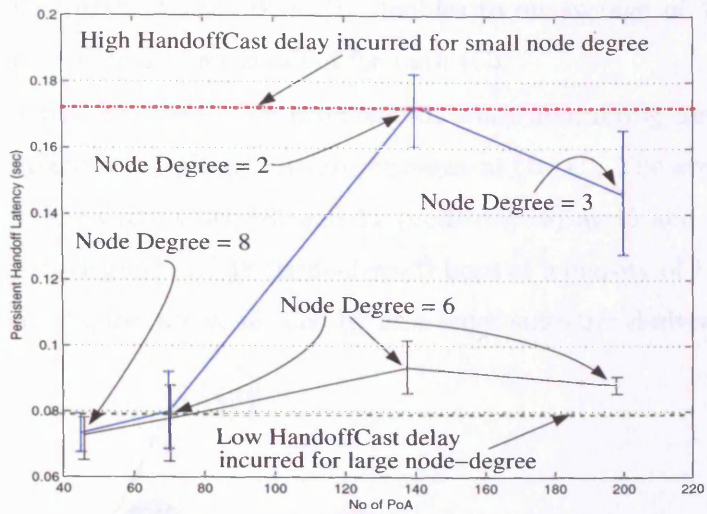


Figure 5.25: 95% Confidence Interval of Persistent Handoff Delay measure effected at previous PoA→RP→new PoA during HandoffCast forwarding, as a function of the number of PoAs.

Figure 5.27 presents the hop measure effected on different segments of the forwarding path between the CN and the MN at both previous and new PoA. It can be seen that *before* MN's handoff, packets arriving at the old PoA experience a nearly constant number of hops; namely, the forwarding segment (A=CN→ old PoA) exhibits a nearly constant average measure of 10-12 hops before MN's IPv6 handoff, depending on the non-leaf AR density of the PoA topology.

On the contrary, *during* MN's handoff, the average number of hops exhibits a measure of variation over the HandoffCast forwarding path (B+C=previous PoA→RP→new PoA). Significant variation on the number of hops may influence the measure of one-way delay component between the previous and new PoA. The variation is *predominantly* dependent on the node-degree of the RP placement. This becomes evident by simply comparing path (A) with each individual segment (B=previous PoA→RP) and (C=RP→new PoA) of the HandoffCast forwarding path (B+C).

The measure of hop variation for path (A) appears to range from 1-2 extra hops, when the non-leaf AR density increases discretely from 45 to 70, 140 and 200 nodes and the shortest path is *not* explicitly effected through the RP. On the contrary, for path (B) and (C) we note that when forwarding is effected through the RP, the average variation in hop count becomes pronounced when the RP node-degree varies significantly; while for a density of 45 (node-deg=8) and 70 (node-deg=6) non-leaf ARs topology, the path-B length does not vary more than 1 hop, at a density of 140 (node-deg=2) and 200 (node-deg=3) the length of path-B nearly doubles to an average of 12 and 10 hops respectively. A similar behaviour is noted for path (C).

The above effect becomes more pronounced, when accounting for the total hop-count of the complete *HandoffCast* forwarding segment (B+C). The average number of hops increases from 12 (node-deg=8) and 11 (node-deg=6) at 45 and 70 non-leaf AR density, to 22 (node-deg=2) and 18 (node-deg=3) hops at a density of 140 and 200 AR, respectively. These figures are confirmed by first order statistics derived in Table 5.2.

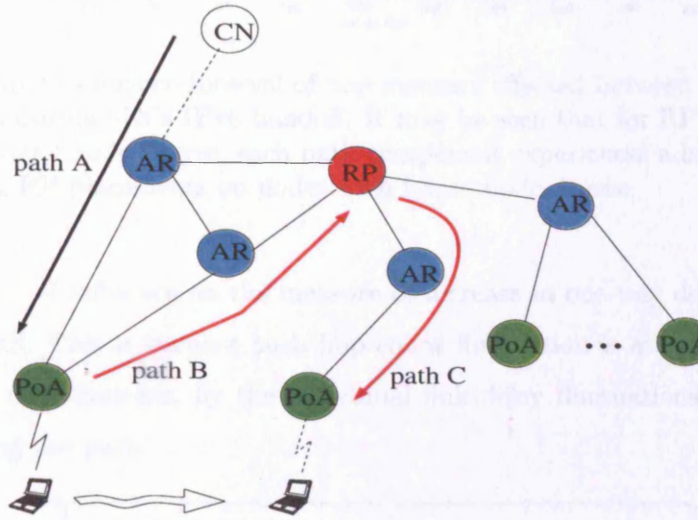


Figure 5.26: Components of one-way delay A=CN→previous PoA, B=previous PoA→RP and C=RP→new PoA comprising the total path between the CN and its MN peer during its IPv6 handoff

The aforementioned variation is magnified when considering the total path length of between the CN and the MN, namely (A+B+C=CN→previous PoA→RP→new PoA); when compared with the original shortest path (A), we notice an increase on the average number of hops per MN handoff, of 10 for PoA densities of 45 and 70 nodes, with 19 or 15 additional hops emerging for PoA densities of 140 and 200 PoAs respectively. The latter attests a significant delay contribution by the additional 9 and 5 hops encountered when the RP placement experiences a small-node degree. This is confirmed by the total one-way delay measure of figure 5.28 tracking the average total CN-MN delay during the course of MN's IPv6 handoff.

The above findings indicate, further that, small variations in the path length (1-2 hops) is insufficient to distinguish between the RP node-degree or PoA density as the

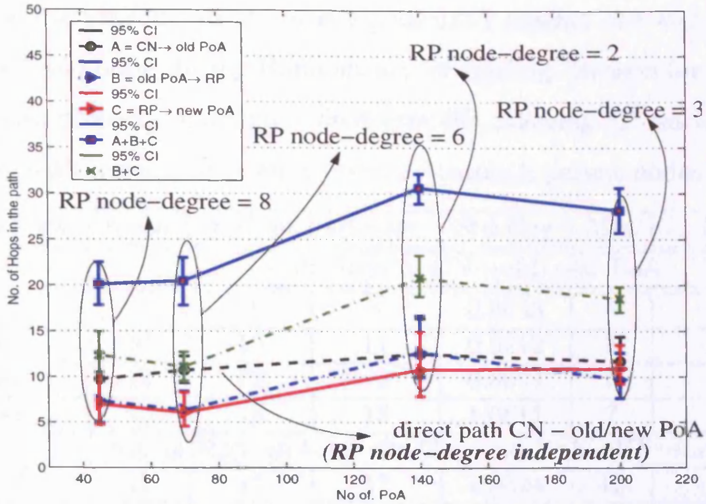


Figure 5.27: 95% Confidence Interval of hop measure effected between the CN and the MN before and during MN's IPv6 handoff. It may be seen that for RP placement onto a node with a small node-degree, each path component experiences a larger increase on hop count than RP placements on nodes with larger node-degree.

dominant source of influence on the measure of increase in one-way delay between the CN and the MN. This is because such hop count fluctuation is absorbed, in terms of one-way delay contributions, by the individual link delay fluctuations experienced by other ARs along the path.

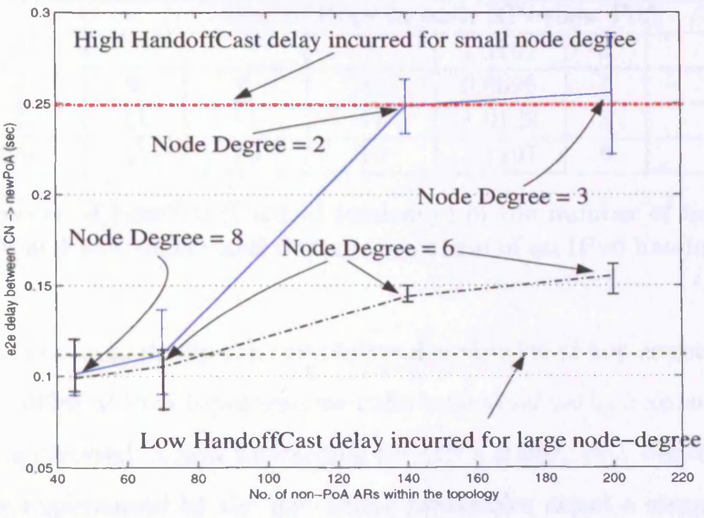


Figure 5.28: 95% confidence interval for total one-way e2e delay for (CN →old PoA→RP→ new PoA) effected during MN's handoff through HandoffCast, as a function of PoA density. Comparing with the previous graph, it appears that the increase in one-way delay and the significant increase in hope count are correlated events.

Thus, the above evidence suggests, that RP placement on a AR node with a higher node-degree (a.k.a fan-in/out) experiences significantly smaller one-way delays over the (old→RP→new PoA) path, during HandoffCast forwarding. Reason for such behaviour appears to be path diversity in a high node-degree RP, allowing SP reachability towards more leaf-PoAs, with infrequent reverse traversal through parent nodes.

PoA density	mean	median	trimean	Std.Dev	Min	Max
#	No. of Hops in CN→old/new PoA					
45	8	8	8	0.9658	5	12
70	10	10	11	0.9212	8	15
140	12	12	12	0.8694	7	14
200	13	13	13	1.0611	7	17
	No. of Hops in total e2e CN→old PoA→RP→new PoA					
45 (8)	18	17	17	1.8524	12	25
70 (6)	21	21	21	1.7045	16	29
140 (2)	31	31	31	1.8128	22	36
200 (3)	28	28	28	1.8730	22	34
	No. of Hops old PoA→RP→new PoA					
45 (8)	12	12	12	1.4682	7	16
70 (6)	11	11	11	1.1953	8	16
140 (2)	22	22	22	1.4148	15	25
200 (3)	18	18	18	1.3897	14	23
	No.of Hops in path old PoA→RP					
45 (8)	7	7	7	1.0696	3	9
70 (6)	6	6	6	0.8692	4	10
140 (2)	12	12	12	1.0002	8	15
200 (3)	10	10	10	0.9779	6	14
	No. of Hops in path RP→new PoA					
45 (8)	7	7	7	1.1101	3	10
70 (6)	6	6	6	0.8695	4	9
140 (2)	11	11	11	1.0129	8	8
200 (3)	10	10	10	1.1197	6	14

Table 5.2: Moments of location (Central tendency) of the number of hops in the path between the CN and MN before and during the course of an IPv6 handoff under HandoffCast.

Table 5.3 shows numerically the cumulative distribution of hop count for the average path length, for different PoA topologies over the individual path components of interest. Where the RP is involved in flow forwarding for MN's traffic, PoA densities signify also the node-degree experienced by the RP. These percentiles attest a steep growth on the average number of hop count, for the majority of the paths used for traffic forwarding either at the HandoffCast path or end-to-end between the CN and the MN. A reduction in the RP node degree by 5, incurs a 69% increase on the number of hops along the path between the CN and the MN, encompassing the HandoffCast path segment. A reduction of 6 in RP's node degree causes a hop-count growth of nearly 88% in the

CN-MN path during an IPv6 handoff.

Looking at the associated measure of one-way delay on the entire path end-to-end between CN-MN, as well as the HandoffCast forwarding path presented in Tables F.3 and F.4 of Annex F, we can confirm that increased hop-count, as a result of a low RP node-degree and the observed measure of one-way delay during MN's handoff are highly correlated; that is to say, the increase in e2e delay between the CN and the MN during an IPv6 handoff, is predominantly due to a low RP node-degree, which in turn increases substantially the hop count of the particular path.

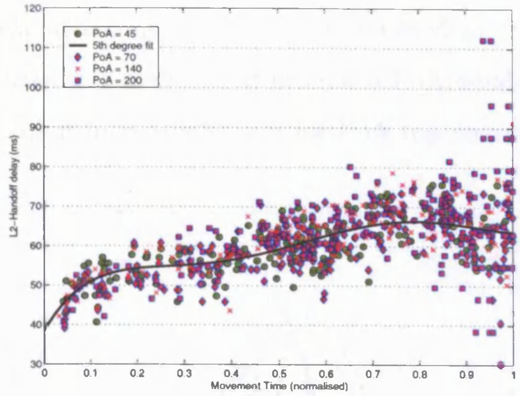
Persistent handoff delay performance, however, encompasses further the measure of L2-handoff delay during an IP handoff. In the following section we analyse the measure of delay contributed by the optimised L2-handoff delay process over 802.11 WLANs, during MN's IPv6 handoff.

PoA density (Node-deg)	Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
#	No. of Hops in CN→ old/new PoA					
45	7	9	9	10	11	12
70	10	11	12	12	13	15
140	11	13	13	13	14	14
200	12	14	14	14	15	17
	No. of Hops in total e2e CN→old PoA→RP→new PoA					
45 (8)	16	18	20	21	23	25
70 (6)	20	22	23	24	25	29
140 (2)	30	32	33	34	35	36
200 (3)	27	29	30	32	34	34
	No. of Hops old PoA→RP→new PoA					
45 (8)	10	12	13	13	15	16
70 (6)	11	13	13	14	15	16
140 (2)	21	23	23	24	25	25
200 (3)	17	18	19	20	21	23
	No.of Hops in path old PoA→RP					
45 (8)	5	7	7	8	9	9
70 (6)	6	7	8	8	9	10
140 (2)	12	13	14	14	14	15
200 (3)	9	11	11	12	13	14
	No. of Hops in path RP→new PoA					
45 (8)	5	7	7	8	9	10
70 (6)	6	7	8	8	9	9
140 (2)	11	13	13	14	14	15
200 (3)	9	10	11	12	13	14

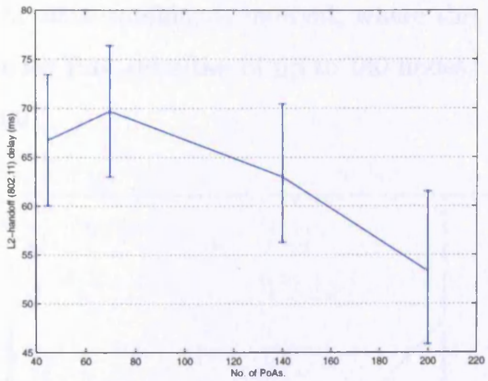
Table 5.3: Percentiles (cumulatively distributed) of the number of hops in the path between the CN and MN before, as well as during the course of an IPv6 handoff under HandoffCast.

L2-handoff Delay contribution

Having looked at the forwarding delay component, we now examine the second delay contribution comprising the total measure of persistent handoff delay. This is the measure of link-layer (L2) handoff delay focused explicitly over 802.11 wireless networks.



(a) L2-handoff delay



(b) L2-handoff delay as a function of PoA densities

Figure 5.29: L2-handoff delay achieved by means or proactive AP probe guiding over different PoA densities. Different PoA densities incur different M-neighbourhood sizes and thus the number of guided channel probes changes per M-neighbourhood

Figure 5.29(a) presents a scatter-plot of the average L2-handoff delay experienced over different PoA densities, during MN's IPv6 handoff. We may see that the measure of L2-handoff delay stabilises around 60-64ms. A larger scatter in the observed delay appears for PoA=200, between 31-110ms, which is confirmed also, by the first order statistics of Tables F.3 and the percentiles of Table F.4 of Annex F. Such scatter is due to a larger variation on the number of operational channels experienced within an R-neighbourhood, over a denser leaf PoA topology; this is shown numerically by the statistics[30] of Table F.5 and the respective percentiles[31] of Table F.6 in Annex F.

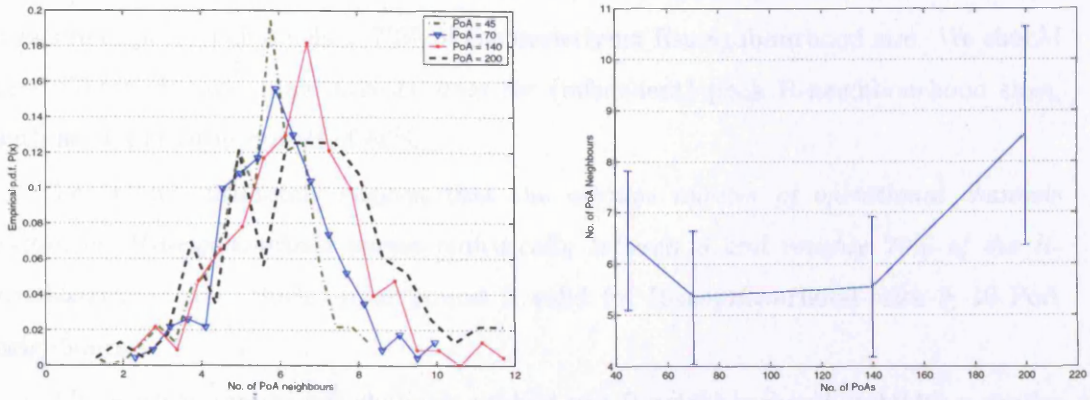
Figure 5.31 presents the empirical probability density function of the average L2-handoff delay, for PoA densities. The larger scatter observed in figure 5.30(a) for PoA=200 is confirmed by heavy tails in the respective density function for that PoA density (figure 5.31-subplot 4).

To appreciate the underlying process of channel probing variation under large PoA topologies, it is important first to look into the average size of MN's R-neighbourhood. Figure 5.30 presents the probability density function of R-neighbourhood size, together

[30] observe difference between mean and median

[31] observed difference between the upper and quartile range

with the number of PoA neighbours as a function of PoA density. From the density function we may see that the number of R-neighbours is normally distributed around 6-7 neighbours independent of the PoA density. For larger PoA densities, the only difference between their respective density functions appears to be the *scale*[32] of the distribution; larger PoA densities exhibit a smaller scale and thus, a larger variance around the mean. This is more evident from the 95% confidence interval, where the variance is maintained around 3 PoA neighbours for PoA densities of up to 140 nodes, while it increases to > 5 for PoA topologies > 200.



(a) Empirical p.d.f P(x) (b) 95% CI

Figure 5.30: Empirical p.d.f and 95% confidence interval on the size of the R-neighbourhood (PoA neighbours) as a function of PoA density

Plotting the number of HAR neighbours as a function of the PoA density provides a clearer measure on both average neighbourhood size as well as error bounds about the mean value for each PoA density. This is shown in figure 5.30(b) where for PoA topologies of up to 140 leaf PoAs, an R-neighbourhood does not exceed the size of 5-8 HAR neighbours. For topologies > 140 the size of R-neighbourhood begins to vary between 7-11 PoA neighbours. The range of error bounds on the particular CI curve indicates that more simulation iterations (i.e. > 20 would be required to attain a more representative performance curve with a smaller measure of variance at the particular confidence level. Despite the measure of error the coarse performance trend indicates that the size of the PoA neighbourhood increases slowly with respect to the PoA density.

Looking at Table F.5 we find that the average size of an R-neighbourhood does *not* imply an equal measure of 802.11 frequency channels operating within that neigh-

[32] A smaller scale in a normal distribution reveals higher variance in the dispersion of the sample values around the mean

bourhood. We find that the actual number of operational channels within that neighbourhood is in fact, *lower* than the size of the R-neighbourhood. Such measure may be expressed as the ratio $\frac{c_n}{s_n}$, where (c_n) is the number of operational channels and s_n the size of the underlying R-neighbourhood.

For neighbourhood sizes between 8-12 PoA, a number of 6-9 operational channels translates roughly to 70-80% of the R-neighbourhood size. Such ratio represents at least the 90-th percentile of the PoA population for each PoA density scenario explored.

From the data of Table F.6 it emerges that, less than 1% of the R-neighbourhoods, with a size of 12, for the maximum PoA density investigated (PoA=200), can sustain a measure of probed channels > 70% of the underlying R-neighbourhood size. We should note that such ratio may increase *only* for (infrequent) peak R-neighbourhood sizes, with an upper ratio bound of 86%.

The above ultimately suggest that the *average number of operational channels within an R-neighbourhood varies realistically between 6 and roughly 70% of the R-neighbourhood size.* Such upper bound is valid for R-neighbourhood sizes $\geq$ 10 PoA neighbours.

The average number of channels probed per R-neighbourhood, exhibits a similar curve pattern to the one of L2-handoff delay, shown in figures 5.29(a) and 5.29(b), with a variation in the number of channels probed, between 4-6. By employing the proactive guiding algorithm during an active AP scan, the L2-handoff process achieves a significant reduction in the number of channels probed by about 61%. Coupled with an optimised `maxChannelTime` period, the total L2-handoff delay is reduced nearly by a factor of 7 from 420ms to about 60-67ms.

Figure 5.29(b) presents the measure of L2-handoff delay as a function of PoA density at a 95% confidence interval. For small PoA topologies (PoA=45,70) the average measure of L2-handoff delay increases marginally from 67 to 69ms. Interestingly, this average delay measure decreases to about 63 and 53ms for an increasing PoA topology (PoA=140,200), with a somewhat larger error bound, than that experienced in smaller PoA topologies. This is due to a higher probe response rate from PoAs ocuppying the same channel, as well as a larger number of PoAs per R-neighbourhood.

From the above analysis on L2-handoff delay performance we may conclude that, the average measure of L2-handoff delay does not exceed 60-70ms for PoA topologies between 45-200 leaf AR nodes. This is achieved by limiting the active AP scanning process over channels operational over MN's R-neighbourhood. The measure of operational

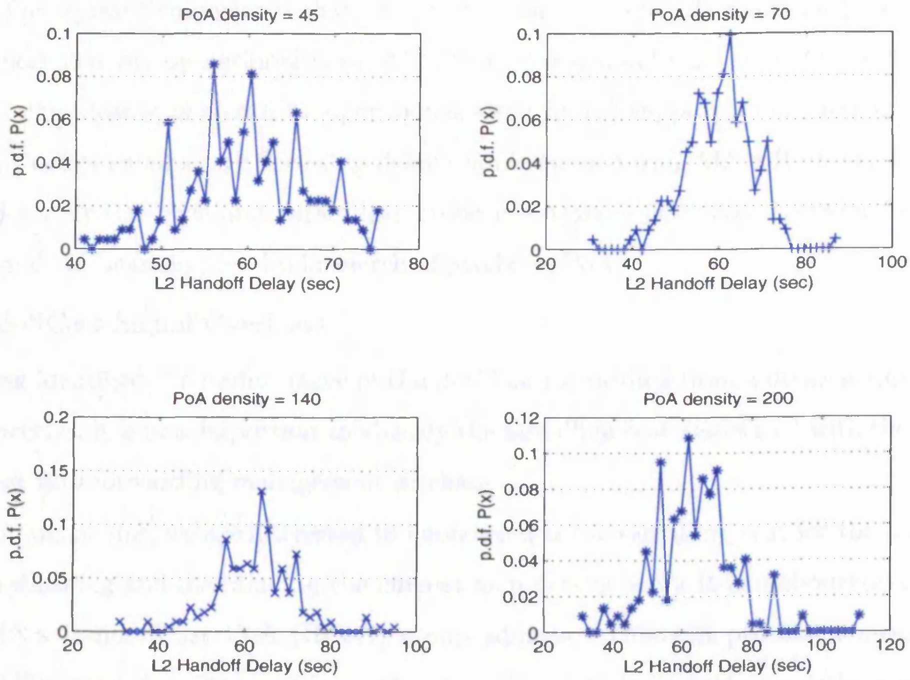


Figure 5.31: Empirical Probability Density Function of L2-handoff delay measure for different PoA densities

channels within an R-neighbourhood is dependent on its size.

Assuming disparity in the WLAN configuration efforts of WISPs, channel allocation over 802.11 PoAs is effectively random. For random allocation of channels among AR members of an R-neighbourhood, we find that the average number of operational channels within an R-neighbourhood is normally distributed with a mean of 6. For an increasing R-neighbourhood size, the variance of the number of operational channels appears to maintain an upper bound of 70% of the R-neighbourhood size (c_n), for $c_n \geq 10$.

For the vast majority of such variation in channel usage, within an R-neighbourhood, the L2-handoff delay remains well within the bound of 100ms. Fewer than 1% of the set of available R-neighbourhoods experienced a number of operational channels > 9. For the L2-handoff process this implies that $< 1\%$ of the monitored L2-handoffs experienced a delay measure $> 100ms$, that did not exceed 112ms. This finding suggests that usage of all 13 available channels is statistically rare with a probability of ≤ 0.005 even for large R-neighbourhoods. Hence, *blind* scanning of all 13 channels during an L2-handoff is concluded to be delay-inefficient, with very low statistical utility.

The above demonstrate that proactive guiding of the AP scanning process over channels that are operational with MN's R-neighbourhood can: (i) reduce L2-handoff delay significantly in a realistic operational scenario, (ii) support a low measure of persistent delay, capable of addressing delay seamlessness during MN's IPv6 handoff, (iii) avoid energy-intensive and packet-loss prone *interleaving* [335, 336] between transmission and AP scanning modes in search of available PoA.

HandoffCast Signal Overhead

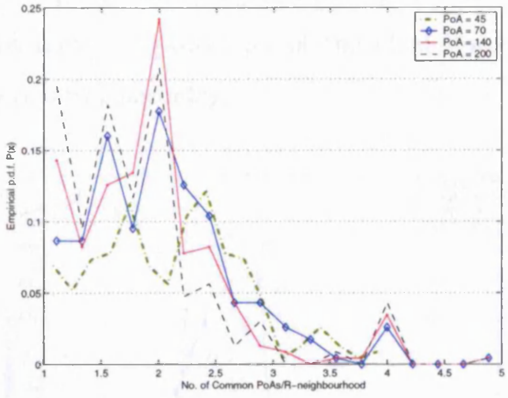
Having identified the performance of HandoffCast forwarding from a delay-seamlessness perspective, it is now important to identify the signalling cost associated with the HandoffCast flow forwarding management mechanism.

In particular, we are interested to know what is the signalling cost for the purposes of establishing and maintaining the current members of MN's R-neighbourhood joined on MN's HandoffCast CoA (HCoA) group address. This will provide a measure of scalability over signalling requirements for significantly large MN populations with an administrative domain.

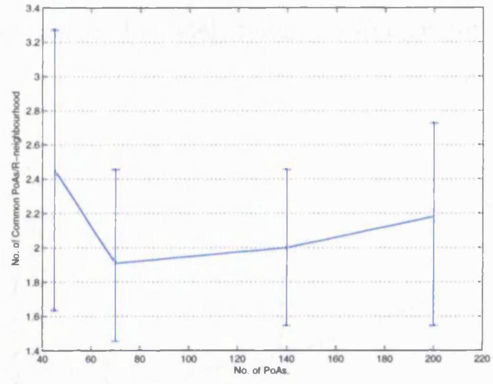
In effort to economise on control signalling, HandoffCast adopts *persistent* messaging, as opposed to *periodic* multicast listener discovery (MLD) reports, employed normally by multicast membership protocols such as IGMP. Intuitively, the measure of saving between periodic and persistent messaging, is dependent on the *interval period* of MLD reports. However, for mobility management purposes, periodic MLD reports should maintain a *frequent* keep-alive signal. This is because, PoAs that do not receive frequently an MLD membership report, can timeout forwarding participation over MN's HCoA group *prematurely* and thus, potentially *reinstate* black-hole effects in the HandoffCast forwarding path between the old and new PoA.

Under persistent HandoffCast signalling, the current PoA would have to effect *two* signals to each PoA that is selected to participate on MN's HCoA forwarding group. Under the periodic signalling counterpart, the number of membership signals would need to be more than two for the duration of MN's cell residence period. Reason for it, is that the current PoA has no knowledge of the duration of MN's cell residence period, to adjust the frequency of membership reports to the minimum possible.

Given that persistent HandoffCast signalling is for mobility management purposes more economic than its periodic counterpart, we now look at the number of signals incurred during the two essential operations in multicast group management abstracted under HandoffCast forwarding management: (i) excluding (a.k.a 'pruning') old invalid



(a) P.d.f



(b) 95% CI

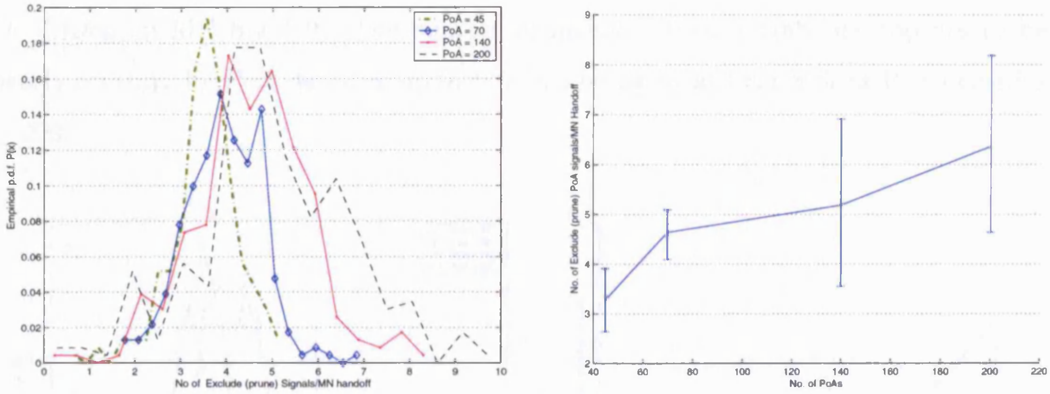
Figure 5.32: Estimating the number of common PoAs *remaining* grafted to MN HCoA group after a handoff, as a function of PoA density

PoA neighbours from MN's HCoA group and (ii) including (a.k.a 'grafting') new valid PoA neighbours onto MN's HCoA group. Common PoA neighbours need not be signalled until they become invalid with respect to MN's current R-neighbourhood.

Figure 5.32(a) presents the probability density function of the number of PoA neighbours, that remain *common* within MN's previous and new R-neighbourhoods during an IP handoff. We may see that for small PoA topologies the probability of common PoA neighbours resembles a shifted-gamma distribution, with mean between 2-3 PoA neighbours. Such measure, appears to decrease slightly for a larger PoA density=70, before it starts increasing very slowly, as shown in figure 5.32(b). This is counter-intuitive, since one would expect to see the number of common PoA increase (even marginally) with leaf-PoA density, as a result of larger R-neighbourhood sizes.

Looking at Annex Table F.5 we note that despite the increase in PoA density, the size of the R-neighbourhood does not increase significantly. In particular, the average size of a PoA neighbourhood for small PoA (density=45,70) topologies, varies between 5-7 neighbours. Such neighbourhood size is representative of at least of 90% of the R-neighbourhood population. For PoA densities=140,200 the size of the R-neighbouhood changes marginally by 1-2 PoA neighbours, as shown in the respective percentiles of Annex Table F.6.

Despite this small increase the number of common neighbours does not increase in denser PoA topologies. The answer to such behaviour is found in the *shape* of the PoA topology. Because during our experiments a higher PoA density is confined within the same pseudo-geographical region[33] the shape of the formed R-neighbourhood does not have the properties of the ideal PoA-neighbourhood model, namely *circular (or hexagon) symmetry.*



(a) p.d.f. (b) 95% CI

Figure 5.33: Estimating the number of PoAs excluded (pruned) from MN's HCoA group after a handoff, as a function of PoA density

As a result of the non-circular asymmetric shape of the R-neighbourhood, appearing mainly in dense PoA topologies (i.e. PoA=140,200), the number of common PoA remains the one observed in smaller R-neighbourhood sizes. Thus, small PoA densities prescribe, the lowest common denominator in *common* PoA neighbours *irrespective* of the PoA density up to 200 PoAs, when such density is confined within the same geographical region.

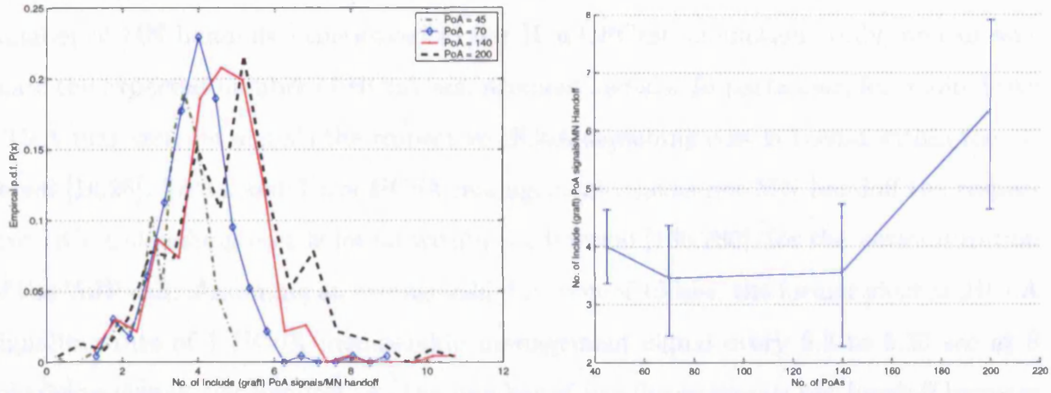
Figure 5.33(b) presents the measure of PoAs excluded from participation in HandoffCast forwarding during MN's IPv6 handoff, by leaving MN's HCoA address. The density function of common PoA neighbours, for all PoA densities appears to be normally distributed with a mean between 3-5 PoAs[34]. As shown in Table F.6 the respective Exclude-PoA percentiles indicate that 90% of the MN handoffs executed experienced between 4-7 leave messages per handoff, which is attested also by the scale of the respective density function of figure 5.33(a). It is interesting to observe that the number of excluded invalid PoA neighbours is monotonically increasing, experiencing larger error bounds (2-3 PoAs) at larger PoA densities. This is, however, expected since the size of the respective R-neighbourhood experiences also similar variation.

Figure 5.34 presents the respective measure of PoAs included to participate in

[33] i.e. the mobility grid - see Section 4.9.3 and Annex E.9.

[34] depending on the PoA density.

HandoffCast forwarding, during MN's IPv6 handoff, by joining MN's HCoA group address. The density function for all PoA densities appears to be normally distributed with an identical mean to the one of the exclude PoA neighbour group. The respective Include-PoA percentiles show an identical 90-th percentile to the one of Exclude-PoA set during an MN handoff. The number of included PoA neighbours appears to be nearly constant for PoA densities up to 140, increasing by at least 50% at PoA densities ≥ 200.



(a) p.d.f. (b) 95% CI

Figure 5.34: The number of PoAs included (grafted) to MN's HCoA group after a handoff, as a function of PoA density

Looking at the average measure of prune and graft signals established on a per MN handoff basis under HandoffCast, we find that each MN's HCoA group address requires on average 6-10 HCoA group membership management signals to sustain an accurate HCoA receiver set during HandoffCast forwarding. For the 90th percentile of all MN handoffs this signalling overhead would increase to 8-14 HCoA membership management signals per MN handoff for the topology densities explored.

For an increasing handoff frequency per MN, the aforementioned measure of signalling becomes more pronounced since the number of signals becomes multiplied by the number of handoffs per unit time.

It is important to remind the reader at this stage that HandoffCast forwarding is effected *only* during the course of IP packet transmissions, such as during the course of a VoIP call. The former implies that a de-association trigger is utilised conditionally by the PoA on the basis of on-going packet forwarding to the MN. From this perspective, an increase on signalling load due to HandoffCast, is effected only for the proportion of *communicating* MNs rather than the entire MN subscriber population.

Furthermore, to assess the scalability of the HandoffCast mechanism under such signalling overhead it is important to consider the average number of handoffs experienced during the course of a voice call. We recall (see Annex F.2), that for an average offered load, depending on the MN speed and the PoA transmission range the number of handoffs per call can essentially vary between 2-20, for the *entire* duration of the (cellular) voice call.

Extrapolating this number of handoffs per VoIP call, on the 50th percentile of the number of MN handoffs experienced in our HandoffCast simulation study, we can estimate the expected number of HCoA management signals. In particular, for 8 join/leave HCoA management signals the respective HCoA signalling cost is bound within the interval [16,28]; for 14 join/leave HCoA management signals per MN handoff the respective HCoA signalling cost is found within the interval [160,280], for the *entire* duration of the VoIP call. Assuming an average call duration of 150sec, the former yield an HCoA signalling rate of 1 HCoA membership management signal every 9.3 to 5.35 sec at 8 join/leave signals per handoff. As the number of join/leave signals per handoff increase to 14, then the HCoA signalling rate changes to 1 HCoA membership management signal every 0.93 to 0.53 sec.

The above implies that for the average case, the *50th percentile* of all MN handoffs the network infrastructure would be required to afford 1 HCoA membership management exclusively over the *wireline* infrastructure every 5-10 sec, depending on the frequency of handoffs. For the *90th percentile* of all MN handoffs, the network infrastructure would be required to afford 1 HCoA membership signal over the wireline infrastructure every 0.5-1 sec depending on the frequency of handoffs. We argue that such signalling cost *is acceptable*, considering that existing reactive Mobile IPv6 management standards prescribe a router advertisement interval of 40ms for the IP-mobility purposes; such a router advertisement interval implies a router advertisement frequency of 1 signal every 0.04 sec over the *wireless* access link. The HCoA membership management cost of HandoffCast under Proactive IPv6 mobility management is at least one order of magnitude smaller than the prescribed MIPv6 interval, in terms of (i) transmission frequency, (ii) control signalling overheads and (iii) bandwidth wastage over the wireless link.

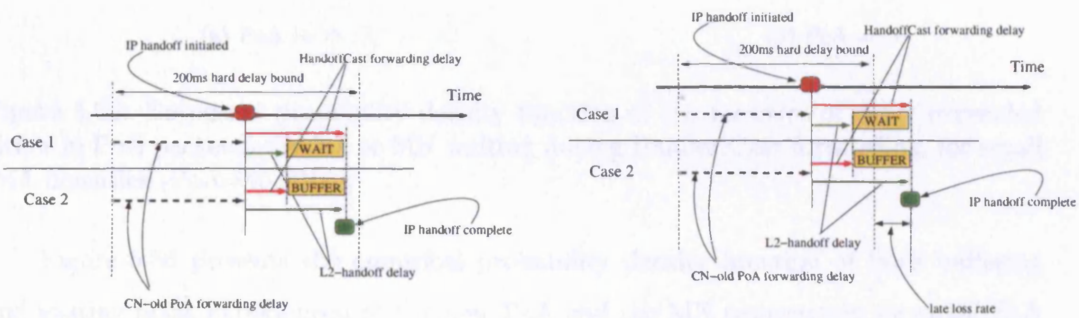
From the above, we may conclude that HandoffCast signalling overheads remains scalable under realistic operational conditions, and perform significantly better than existing reactive MIPv6, in terms of signalling requirements. Such performance increase arises from (i) robust HandoffCast forwarding and reduction/elimination of black hole effects, typical on reactive MIPV6 standard solutions (ii) significantly smaller HandoffCast signalling overheads compared to router advertisement overheads prescribed by MIPv6 standards.

Buffering/Waiting Delay

Having identified the performance of forwarding and L2-handoff delay, it is important to also look at the measure of persistent handoff delay that is transformed onto buffering of MN's traffic, during an IP handoff.

Before analysing buffering performance under HandoffCast, it is important to note that, during an IPv6 handoff, an *L2-handoff* and *HandoffCast forwarding* are initiated *concurrently,* as the MN de-associates with its previous PoA.

Thus, forwarding delay performance *during* an IP handoff is tracked by the *slowest* of the above two functions. The emerging issue is to identify which of the two processes appears to be the slowest.

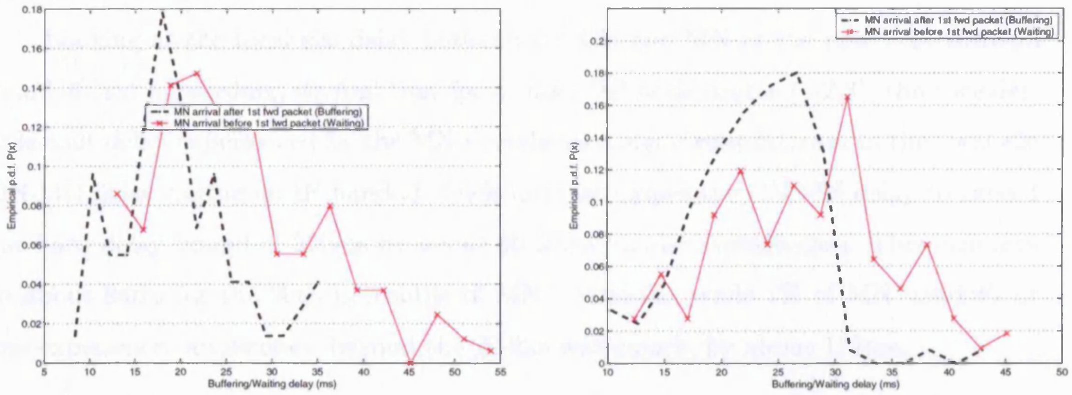


(a) Buffering within the RT hard deadline (b) Buffering exceeding the RT hard deadline

Figure 5.35: Buffering and Waiting delay depending on the difference in measure between the L2-handoff and HandoffCast forwarding delay. Depending on whether the total e2e path between CN and MN during an IPv6 handoff exceeds the 200ms delay bound packets arriving may unavoidably face a short late loss rate.

Figure 5.35(a) illustrates the two distinct cases, where: (i) the measure of L2-handoff delay is smaller than the HandoffCast forwarding delay (ii) the measure of HandoffCast forwarding delay is smaller than the L2-handoff delay. In the first case, the MN appears to arrive *early* at the new PoA link in comparison to the first HandoffCast-forwarded packet. As a result, the MN *waits* the arrival of packets at the new PoA. In the second case the MN appears to arrive *late* at the new PoA link compared to the first HandoffCast-forwarded packet. In such event, HandoffCast-forwarded packets are *buffered* at the new PoA, until MN's re-association triggers the conditional forwarding of the buffer content to the attached MN.

In both of these two cases the e2e delay between the CN and the new PoA when the slowest of the two functions completes, *remains* below the 200ms hard delay bound for interactive real-time services. Figure 5.35(b) illustrates the case where the e2e delay between the CN and the new PoA, when the slowest of the two functions completes, *exceeds* the 200ms hard delay bound.



(a) PoA = 45 (b) PoA = 70

Figure 5.36: Empirical probability density function of the measure of delay expended either in PoA packet buffering or MN waiting during HandoffCast forwarding, for small PoA densities (PoA=45,70)

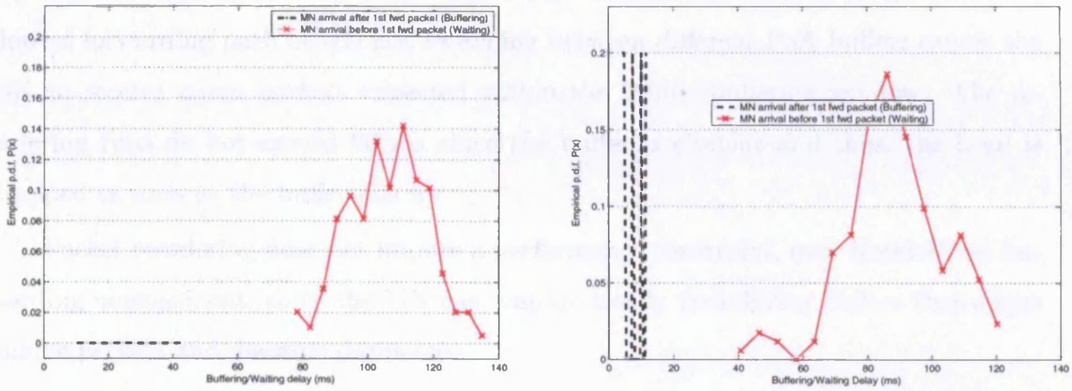
Figure 5.36 presents the empirical probability density function of both buffering and waiting delay experienced at the new PoA and the MN respectively for small PoA densities. It may be seen that for a PoA topology of 45 ARs the probability of PoA buffering versus MN waiting is only marginally larger by about 4%. Such difference is eliminated for a PoA density=70; both events appear to be equiprobable for that topology size.

The above is confirmed by the respective percentiles of Table F.4; at the 90th percentile, we see that despite the somewhat greater probability of PoA buffering during MN's IP handoff, at PoA=45 (ND=8), the MN appears to spend only marginally more time in waiting than the PoA spends in buffering by approximately 10ms. Such difference reduces further to 5ms when the RP node-degree reduces by 2 for PoA=70 (ND=6).

For large PoA densities where the RP node-degree is small, we observe that the measure of PoA buffering is negligible compared to the measure of MN waiting. For node=degree=2 the MN appears to wait for the first HandoffCast-forwarded packet between 92-120ms, having arrived early at the new PoA due to a comparatively faster L2-handoff.

The above indicate that under HandoffCast forwarding management for a *small* RP node-degree, the MN *is more probable to arrive early* than late at a new PoA. This in turn implies *larger* waiting times and *little or no* PoA buffering. On the contrary, for a *large* RP node-degree, the probability of an *early or late* MN arrival at the new PoA is nearly equal.

Looking at the total e2e delay between the CN and MN at the new PoA through HandoffCast forwarding, we find that for a *small* RP node degree (=2,3), the considerable wait delay experienced by the MN introduces a *significant increase* in the total e2e CN-MN delay during an IP handoff. Such increase causes the CN-MN delay to exceed the hard delay bound of 200ms by about 30-50ms for the average case. This increases to about 84ms for the 90th percentile of MN's handoffs, while 1% of MN handoffs or less experiences an increase beyond the 200ms watermark, by about 122ms.



(a) PoA = 140 (b) PoA = 200

Figure 5.37: Empirical probability density function of the measure of delay expended either in PoA packet buffering or MN waiting during HandoffCast forwarding, for large PoA densities (PoA=140,200)

On the contrary for a *large* RP node degree (=6,8), the e2e delay between the CN-MN path at the new PoA during an IP handoff appears to be *well below* the hard delay bound of 200ms; in particular, for an RP node degree=8 the average e2e delay between CN and MN at the new PoA through HandoffCast forwarding is 98ms. For an RP node degree of 6 the average e2e delay increases by an extra 20ms (118ms). For the 95th percentile of MN handoffs over HandoffCast forwarding employing large RP node degrees, the e2e delay experienced is 126ms (ND=8) and 146ms (ND=6) respectively. This is clearly within the delay performance bounds targeted by HandoffCast forwarding performance.

Ultimately, we may note that where PoA buffering is employed, the MN does not experience packet reordering for more than 100ms, in the extreme case of ping-pong effects and 40-50ms under normal IP handoffs. Such measure of packet reordering is significantly *smaller* than the one encountered in the M&M micro-mobility management protocol proposed by Helmy et al [232], which also employs multicast-based forwarding, with the RP placed at the border router. In this proposal the authors report a packet reordering period of 500-600ms, which in turn suggests a rather *slow-reacting* forwarding management proposal, compared to the proposed HandoffCast function.

Packet re-ordering may only occur when the HandoffCast forwarding delay between the old PoA and each candidate PoA neighbour is significantly different. As the MN handoffs fast at PoA1 packets are forwarded to the MN; however, in the event of a ping-pong (or other event that causes a very high instantaneous handoff rate) the MN oscillates its handoff between two or more PoA neighbours. At that stage either the slowest forwarding path or the fast switching between different PoA buffers causes the MN to receive *again* packets expected within the 200ms buffering window. The re-ordering runs do not exceed 200ms since the buffer is circular and thus, the head is dropped as soon as the buffer fills up.

Packet reordering does not impose a performance constraint, over HandofCast forwarding management, since the MN can employ locally re-ordering buffers that aligns unique packets and discards duplicates.

5.7 Conclusions

This chapter has investigated the performance of HandoffCast as a proactive flow forwarding mechanism. HandoffCast is geared to complement the overall Proactive IPv6 mobility management architecture towards support of delay seamlessness. Such advanced network-layer capability is targeting robust support of real-time interactive services deployed over next generation wireless IPv6 Access networks.

To the best of our knowledge this work is the first in its kind addressing systematically flow forwarding management for IPv6 mobility purposes. To this end, there exists little experimental evidence that may provide a comparative or validating measure against the derived measure of delay performance. Necessary and sufficient criteria employed to ensure validity of the derived simulation results, have been the accuracy of network topology generation and RTT distribution within the generated network topology, all driven by actual experimental measurement results reported in the literature.

From a detailed set of simulations conducted to assess the coupling of HandoffCast performance with Proactive handoff management, we arrived at a number of important results that act in support of the experimental hypothesis: proactive forwarding management through HandoffCast *is* capable of addressing successfully delay seamlessness over wireless access networks. Such capability is, however, directly related to specific performance factors, identified and optimised throughout this simulation study.

These factors stem from two fundamental delay sources that remain beyond the explicit control of the network layer: (i) the measure of link-layer (L2) handoff effected by the individual wireless technology at hand, (ii) the measure of one-way delay arising from the perspective of flow forwarding during MN's IP handoff, in an effort to eliminate black-hole effect and reduce the measure of perceived flow disruption.

The above two factors have been collectively identified as the measure of *persistent handoff delay*, since both L2-handoff and one-way delay during flow forwarding *persist* to some minimum measure, during MN's IP handoff. With respect to the L2-handoff process, our investigation has focused specifically over IEEE 802.11 WLAN networks, since cellular L2-handoff performance is well within the delay bound of 100ms targeted by HandoffCast performance requirements.

5.7.1 Assessing explicitly L2-handoff delay optimisations

We have identified that HandoffCast performance is critically dependent on the measure of the delay contributed by the standard 802.11 L2-handoff process. To this end, we have proposed an L2-handoff delay optimisation that aims to reduce drastically the latency incurred during the critical period of active PoA scanning with no negative impact to function robustness.

Such optimisation operates by means of (i) proactively guiding the AP scanning phase to probes over channels that operate *explicitly* over MN's current PoA neighbourhood, (ii) reducing the maximum dwell time (`maxChannelTime`) per probed channel. With respect to state pertinent to proactive guiding of the AP scanning process, such information is readily available during HAR discovery.

With respect to the measure of the maximum dwell time per probed channel, we have shown, through simulation, that such optimisation is feasible by considering a trade-off between the expected *peak AP capacity*[35], while *increasing the PoA density* for mobility management purposes. The combined effect of the two L2-optimisations guided by network-layer triggers, was found to be capable of providing the targeted L2-handoff delay reductions and was thus, employed in subsequent HandoffCast performance analysis.

5.7.2 HandoffCast forwarding performance

With respect to HandoffCast forwarding, its performance in terms of bounded delay, below the 200ms delay bound, appears to be satisfactory but dependent on the placement of the RP. Our simulation results suggest that HandoffCast performs well for an RP node-degree ≥ 6. On the contrary, for an RP node-degree of ≤ 3, the measure of persistent delay grows significantly on the forwarding path (old PoA $\rightarrow$ RP $\rightarrow$ new PoA). Such effect is exacerbated when considering the total e2e delay over leaf-PoA densities ≥ 140, between the CN $\rightarrow$ old PoA $\rightarrow$ RP $\rightarrow$ new PoA. There, the total one-way delay is found to be in excess of 200ms, with an average e2e delay value of 255ms and peaks around 300ms. This implies that inappropriate placement of the RP, namely, onto an AR with a *low* node-degree (≤ 3) renders HandoffCast forwarding delay prohibitively large for growing leaf PoA topologies ≥ 140.

On the contrary, RP placement at an AR with a *high* node-degree (≥ 6) sustains *scalable* HandoffCast forwarding delay performance for increasing PoA densities. Our results suggest that by placing, instead, the RP on a high node-degree AR, even on topologies with a large leaf-PoA density, the HandoffCast path delay on the HandoffCast forwarding path (old PoA $\rightarrow$ RP $\rightarrow$ new PoA), remains near the delay measure observed for smaller PoA densities.

5.7.3 L2-handoff delay performance under HandoffCast

With respect to L2-handoff delay performance during HandoffCast forwarding, we have observed an average measure of L2-handoff delay that does not exceed 60-70ms for small or large PoA densities (i.e. 45-200 leaf-PoA nodes).

Given a random allocation of channels among members of an R-neighbourhood that comprises of different WISPs, we find that the average number of operational channels within an R-neighbourhood is normally distributed with a mean of 6. For an increasing R-neighbourhood size, the variance of the number of operational channels appears to maintain an upper bound of 70% of the R-neighbourhood size (c_n), for $c_n \geq 10$.

[35]In the form of number of wireless stations or traffic load.

Fewer than 1% of the monitored L2-handoffs experienced a delay measure $> 100ms$, that did not exceed 112ms, making such L2-handoff delays a rare event.

The above results confirm that proactive guiding of the AP scanning process over channels that are operational with MN's R-neighbourhood can: (i) reduce L2-handoff delay significantly in a realistic operational scenario, (ii) support a low measure of persistent delay, capable of addressing delay seamlessness during MN's IPv6 handoff, (iii) avoid energy-intensive and packet-loss prone *interleaving* [335, 336] between transmission and AP scanning modes in search of available PoA.

5.7.4 HandoffCast Signalling Overheads

With respect to the measure of signalling overheads, we find that each MN's HCoA group address requires on average 6-10 HCoA group membership management signals to sustain an accurate HCoA receiver set during HandoffCast forwarding. For the 90th percentile of all MN handoffs this signalling overhead would increase to 8-14 HCoA membership management signals per MN handoff for the topology densities explored.

The above implies that for the average case *50th percentile* of all MN handoffs the network infrastructure would be required to afford 1 HCoA membership management exclusively over the *wireline* infrastructure every 5-10 sec depending on the frequency of handoffs. For the *90th percentile* of all MN handoffs, the network infrastructure would be required to afford 1 HCoA membership signal over the wireline infrastructure every 0.5-1 sec depending on the frequency of handoffs.

We have argued that the above signalling cost *is acceptable*, considering that existing reactive Mobile IPv6 management standards prescribe control signalling pertinent to its operation (router advertisements), at an interval of 40ms; this implies a router advertisement frequency of 1 signal every 0.04 sec over the *wireless* access link. The HCoA membership management cost of HandoffCast under Proactive IPv6 mobility management is at least one order of magnitude smaller than the prescribed MIPv6 interval, in terms of (i) signalling overheads and (ii) bandwidth wastage over the wireless link.

From the above we have concluded that HandoffCast signalling overheads remains scalable under realistic operational conditions, performing significantly better than existing reactive MIPv6 signalling requirements. Such performance increase arises from (i) robust HandoffCast forwarding and reduction/elimination of black hole effects, typical on reactive MIPV6 standard solutions (ii) significantly smaller HandoffCast signalling overheads compared to router advertisement overheads prescribed by MIPv6 standards.

5.7.5 HandoffCast Buffering/Waiting delays

In regards to the measure of buffering required at a new PoA, we have looked at the measure of L2-handoff and HandoffCast forwarding delays. Given that an L2-handoff and HandoffCast are concurrent events, we have assessed the conditions under which the slowest of the two events tracks either: (i) the measure of PoA buffering or (ii) MN waiting.

We have found that during HandoffCast forwarding, a *small* RP node-degree, induces statistically *an early MN arrival* at a new PoA, compared to the arrival of the first HandoffCast forwarded packet. This implies *larger* waiting times and *little or no* PoA buffering. By contrast, for a *large* RP node-degree, *early or late* MN arrival at the new PoA are equiprobable events, giving rise to a uniform probability between PoA buffering or MN waiting.

Looking at the total e2e delay between the CN and MN at the new PoA through HandoffCast forwarding, we find that for a *small* RP node degree (=2,3), the considerable wait delay experienced by the MN introduces a *significant increase* in the total e2e CN-MN delay during an IP handoff. Such increase causes the CN-MN delay to exceed the hard delay bound of 200ms by about 30-50ms for the average case. This increases to about 84ms for the 90th percentile of MN's handoffs, while 1% of MN handoffs or less experiences an increase beyond the 200ms watermark, by about 122ms. Based on these results, we have concluded that for a small RP node degree, buffering is not applicable since the wait delay of the MN, renders all arriving packets a late loss rate for the duration of the handoff.

On the contrary for a *large* RP node degree (=6,8), the e2e delay between the CN-MN path at the new PoA during an IP handoff appears to be *well below* the hard delay bound of 200ms. For an RP node degree=8 the average e2e delay between CN and MN at the new PoA through HandoffCast forwarding is 98ms. For a 60-70ms L2-handoff delay this implies an MN wait delay (i.e. early arrival) with respect to the first HandoffCast forwarded packet of 20-50ms.

For an RP node degree of 6 the average e2e delay increases by an extra 20ms (118ms). For the 95th percentile of MN handoffs over HandoffCast forwarding employing large RP node degrees, the e2e delay experienced is 126ms (ND=8) and 146ms (ND=6) respectively. This implies an MN wait delay of 50-70ms, suggesting that during HandoffCast forwarding, no buffering occurs for an L2-handoff delay up to 70ms. From that we may conclude that while flow forwarding delay sustains a one-way delay between the CN and the new PoA of $< 200ms$, the early arrival of the MN at the new link, tends to remove statistically the need for buffering while the L2-handoff delay remains smaller than the HandoffCast forwarding delay.

For 99% of the flows forwarded through an RP with a high node degree, the buffering requirement at the new PoA did not exceed the measure of 50ms. Such delay excess accounts effectively for the possibility that the L2-handoff is *slower* than the HandoffCast forwarding delay by an equal delay measure. This expected to occur only on very high PoA densities ($>> 200PoAs$) with R-neighbourhood size > 15. Hence, under HandoffCast forwarding and for a PoA density up to 200, the measure of packet buffering does not exceed 50ms, before MN's IPv6 handoff is complete.

Chapter 6

Proactive MIPv6 vs Fast Handoffs for MIPv6

Latency reduction during a handoff, from the perspective of frequent change of IPv6 PoA, has been also the objective of several mobility management mechanisms aiming to extend Mobile IPv6 [301, 337, 338].

These IPv6 mobility management extensions have also been adopted as workgroup items in the MIPv6 Signalling and Handoff Optimisation (MIPSHOP)[1] WG of the IETF [193, 339, 301, 338]. Amongst them, Fast Handoffs for MIPv6 [340], is emerging as the dominant work-group draft specification of IPv6 mobility management optimisations, targeting to support frequent change of PoA for the MN on the move.

In this section we present a comparative analysis between the proactive IPv6 mobility management architecture and the latest[2] Fast Handoffs workgroup draft recommendation [33]. For brevity, Proactive IPv6 Mobility management is identified as PMIPv6 and Fast Handoffs as FMIPv6; these terms will be used interchangeably, where necessary, throughout this analysis.

To provide an objective comparative analysis over performance and functionality we present briefly a functional evolution of the two proposals. This facilitates an assessment also as a measure of the functions employed, given that FMIPv6 has been subject to continuous development and thus, change. The latter makes a performance analysis between PMIPv6 and FMIPv6 difficult to track or compare, while retaining a realistic comparative basis, drawn from earlier reported performance [337]. It is for this reason that we provide a functional perspective of the two proposals, which ultimately comes to justify certain design trade-offs in each approach.

---

[1]This IETF working group has branched-off the MOBILEIP WG during the last 1-2 years.
[2]At the time of writing its version is captured by the respective reference.

6.1 Functional Perspective between PMIPv6 and FMIPv6

The PMIPv6 architecture was first proposed as an independent study towards seamless mobility management, in [341], as well as an individual draft recommendation to both MOBILEIP and SEAMOBY WGs in [129]. At the same time the Fast handoffs design team, established the basis of the FMIPv6 draft recommendation in [342].

A number of architectural and functional differences separated the two mobility management proposals at the time; these are summarised in table 6.1.

MM functions	PMIPv6	FMIPv6
mobility-hop routing	√	×
CAR discovery	√	×
M-R neighbourhood mapping	√	×
context-state establishment	√	×
MN flow forwarding (unicast)	×	√
MN flow forwarding (multicast)	√	×
Lossless Ping Pong management	√	×
assumes accurate NAR determination	×	√
Depends on Mobile IPv6	√	√

Table 6.1: Key functional differences between PMIPv6 and early revisions of FMIPv6.

Proactive MIPv6 [47, 49, 50, 52] has remained in most part unchanged since then. The FMIPv6 specification has undergone a number of major revisions [343, 340, 344, 345, 346, 347, 348, 349, 350, 33]; these revisions have gradually overloaded the semantics of the original FMIPv6 signalling (e.g. HI/HAck), by adopting key functions of the PMIPv6 architecture. An example of this is the M-R neighbourhood mapping in the form of an [AP-ID, AR-Info] tuple in FMIPv6, as specified in [345].

Proactive MIPv6 mechanism	Proactive MIPv6	Fast MIPv6
M-R Neighbourhood Discovery	ver.00 (7/2001)	-
mobility-hop Routing	ver.00 (7/2001)	-
M-R neighbourhood mapping	ver. 00 (7/2001)	ver.06 (3/2003)
Handoff Capability	ver.00 (7/2001)	ver.03 (12/2001)
Context Capability	ver.00 (7/2001)	ver.07 (7/2003)
proactivity/predictive	ver.00 (7/2001)	ver.03 (12/2001)
HandoffCast (multicast)	ver.00 (7/2001)	-

Table 6.2: Key PMIPv6 mechanisms influencing subsequent revisions of FMIPv6.

Table 6.2 provides a table of key functions found at the core of the PMIPv6 architecture [341, 129], and its adoption by subsequent revisions of the FMIPv6 draft specification.

The influence of the PMIPv6 architecture on seamless mobility specification is evident also over the mechanisms of Handoff AR discovery. Based on the work of [50] and the original protocol specification work in [129], Chalmers et al [147] and Trossen et al [351] have standardised issues arising in candidate access router discovery, giving rise to a recent candidate access router discovery draft specification, identified as `DyCard` [352] and fairly recently `CARD` [353] standard. It may be noted that our original proposal for M-R Neighbourhood discovery, encompasses a superset of the algorithms implemented in the DyCard draft specification.

Despite the positive contribution of architectural features of PMIPv6 over the FMIPv6 specification, our study finds that the latest FMIPv6 specification is presented with significant issues related to handoff delay performance towards interactive real-time services.

In the following sections, we analyse signalling and trade-offs between the two mobility management mechanisms. We present strengths as well as performance limitations of the FMIPv6 recommendation with respect to handoff delay optimisations and contrast it against the performance of PMIPv6.

To the best of our knowledge, PMIPv6 is the first tightly-coupled protocol architecture aiming to address delay seamlessness at both the network (IPv6) and (for delay-prone wireless technologies) link layer in a systematic way. Such architecture identifies clearly performance gains through *explicit* link-layer triggers, while addressing architecturally a multitude of mobility management signalling aspects pertaining to delay seamlessness for interactive IP services.

6.2 FMIPv6 vs Proactive MIPv6 performance

This section analyses functional aspects between FMIPv6 and PMIPv6 and identifies trade-offs and expected measure of handoff performance for the two mechanisms. While the set of technical issues identified in this analysis is not exhaustive[3], its identifies significant issues in the design of the latest FMIPv6 specification revision. We feel these must be addressed before the recommendation matures to an experimental standard.

6.2.1 NCoA determination and tunnel configuration efficiency

This section elaborates on performance issues arising under FMIPv6 during the process of NCoA determination, upon detection of new APs[4] by the MN, over its wireless interface. Such issues are contrasted with the performance of PMIPv6 under similar

[3]the exhaustive set of performance issues identified in our FMIPv6 comparative study is omitted in this analysis for brevity.

[4]through their respective AP identifiers, e.g. BSSIDs in 802.11b/g

operational conditions.

Section 6.2.5 has elaborated on the performance of the abstraction[5] of link-specific methods to support detection of new APs, by the MN while in active communications with its peers. We have seen that functions such as interleaving between AP scanning and packet transmissions, appear to *impact negatively* on the performance of *any* on-going interactive real-time flow between the MN and its peers. This has been shown also through experimental evidence [180, 354] particularly when the MN has to contend with other wireless stations for medium access.

However, for the purpose of elaboration we assume the hypothetical possibility that the MN can, by some yet to be defined, link-specific methods, identify new APs in its transit path, while: (i) in real-time packet communications with its peers (ii) in advance of a link-layer handoff.

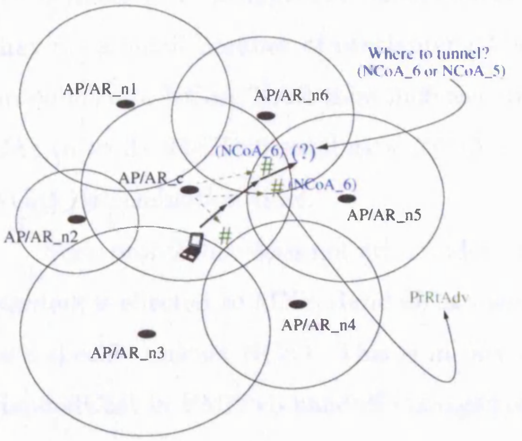
According to the FMIPv6 draft specification, resolution of an AP-identifier to subnet-specific information, aims to allow an MN to configure a prospective new CoA (NCoA). To this end, an AP-ID identifier sent to PAR[6] through an RtSolPr results in a PAR response specified by the FMIPv6 specification as follows:

Specification 6.1 *'..In response the Access Router (PAR) sends a PrRtAdv message which contains one or more [AP-ID, AR-Info] tuples. ..' (pg.6-7)*

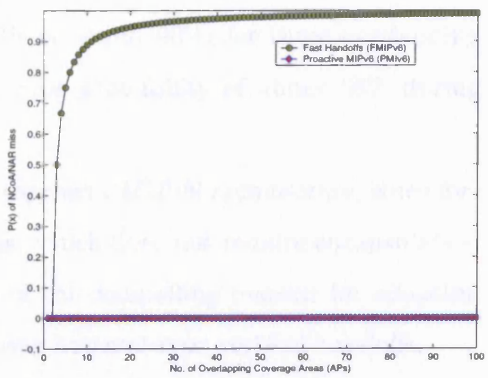
However, in the hypothetical[7] situation that multiple AP-identifier information is available, *the resolution of multiple AP identifiers to AR information does not provide any assurance that a particular PoA (AP/AR) is a primary candidate for an IPv6 handoff.*

This is illustrated in figure 6.1(a). As the MN moves away from AP/AR_c[8] we assume (according to the FMIPv6 specification) it identifies by means of link-specific methods, two new APs: AP/AR_{n6} and AP/AR_{n5}. For each of these APs the PAR responds with a PrRtAdv with subnet-specific information of their respective ARs, such that the MN generates with equal probability two NCoAs: $NCoA_6$ and $NCoA_5$. However, as the MN departs unequivocally from the boundaries of AP/AR_c, the MN is required (see Spec.6.6) to instruct the PAR to bind MN's previous CoA PCoA) to NCoA. Since there exists more than one prospective NCoAs configured ($NCoA_6$ and $NCoA_5$), FMIPv6 fails to determine how the MN identifies which NCoA to include in

[5]no such methods have been either identified, referred to, or specified by the FMIPv6 specification.
[6]termed as AR_c under PMIPv6.
[7]In 802.11 WLAN networks, the MN can associate typically with only a single AP at any time
[8]identified as PAR in FMIPv6



(a) NCoA determination uncertainty in FMIPv6



(b) NCoA/NAR miss Probability

Figure 6.1: NCoA determination uncertainty under FMIPv6 with a cascading effect on FBU signalling sent to PAR. Associated probability of NCoA/NAR miss for an increasing number of overlapping CAs

its FBU to the PAR, for tunnel setup purposes. It can be seen that under such (not uncommon) conditions, FMIPv6 faces a probability of 0.5 of erroneous tunnel setup; we identify this as *NCoA/NAR miss* probability.

It is noted that a NCoA/NAR miss during FBU transmission time, will cause the PAR to forward MN's traffic to the wrong PoA, hence re-introducing packet loss due to forwarding. This gives rise to *migrant black-hole effects* over the VoIP flows between the MN and its peers.

In fact, we deduce that the probability of a NCoA/NAR miss is proportional to the number of overlapping coverage areas (CAs) n within a single AP-cell; it is equal to $P(x) = 1 - \frac{1}{n-1}$. Such is the case with *vertical handoffs*, whereby multiple WISPs supporting the *same* wireless technology (such as 802.11b/g networks), overlay *vertically* over the same geographical location. The above indicate that the targeted performance of the FMIPv6 specification remains severely limited in cases of multi-ISP wireless overlay network deployment.

The issue of increased NCoA/NAR probability arises *also* under FMIPv6 for the common case of *horizontal handoffs* in overlap regions of typical M-neighbourhood configurations with six or more immediate PoA neighbours. In such M-neighbourhood configurations, the most frequent case of multi-cell overlap within a CA is the one of a 3-cell overlap region, shown also in figure 6.1(a). Figure 6.1(b) shows the probability of NCoA/NAR miss for an increasing number of *overlapping coverage areas (CAs)* within the boundaries of a single cell, independent of the form[9] of IP handoff. It can be seen that for a small number of overlapping CAs ($n \leq 10$) within an overlap region, the probability of NCoA/NAR miss increases rapidly to about 90%; for three overlapping CAs ($n = 3$), FMIPv6 exhibits a NCoA/NAR miss probability of about 68% during NCoA determination time.

Such probability does not arise under the proposed PMIPv6 architecture, since forwarding is effected to MN's HandoffCast address, which does not require encapsulation to a specific unicast NCoA. This is in fact one of the compelling reasons for adopting HandoffCast in PMIPv6 handoff management over horizontal or vertical handoffs.

6.2.2 Frequency of NCoA/NAR miss probability for FMIPv6

Besides the number of overlapping cells over a particular overlap region, the measure of NCoA/NAR miss probability is also dependent on the frequency of occurrence *within* a CA, as well as the measure of area overlap between neighbouring CAs. To this end, we analyse the behaviour of overlap between multiple CAs and demonstrate its impact onto the NCoA/NAR miss probability from FMIPv6. This analysis focuses on the measure of NCoA/NAR miss probability of FMIPv6 as a function of *inter-AP distance* d_i and *CA radius* r_i.

For the purposes of this analysis, CA radius is the defined as the measure of transmission range of an AP, assuming a circular coverage area pattern; the measure of inter-AP distance d between two APs is defined as:

- *small*, if at least one AP falls within the transmission range of the other (i.e. $d \leq max(r_0, r_1)$).
- *large*, if no AP falls within the transmission range of each other with no area overlap over their respective CAs (i.e. $d = r_0 + r_1$).
- *normal*, if no AP falls within the transmission range of each other while the measure of area overlap for their respective CAs is tracked by $max(r_0, r_1) \leq d \leq r_0 + r_1$.

We analyse the behaviour of NCoA/NAR miss probability by means of devising a synthetic three-cell overlap model emerging from the clustering of multiple two-cell overlap regions comprising a 6/7 cell M-neighbourhood pattern. The model can be generalised for n-cell overlap, by incorporating the Edelsbrunner inclusion-exclusion

[9] horizontal or vertical

formula [355]; the latter is applicable onto the simplicial complexes[10] emerging within the overlap area. We, however, postpone the analytic modelling of n-cell overlap as future work extension, since it falls beyond the scope of this analysis.

Annex G.1 presents the parametric derivation of a basic two-cell overlap model, extended subsequently into a three-cell overlap system explored through simulations.

Devising a 3-cell overlap model

The 2-cell overlap model is augmented with a third coverage area. To model the effect of transmission range on the measure of overlap, we assume that the two existing overlapping CAs (CA_A) and (CA_B) have the same radius, i.e. ($r_0 = r_1 = R$). The third coverage area (CA_K) is then represented with centre K($x_2, y_2 + 2y + r$) with radius r as shown in figure 6.2.

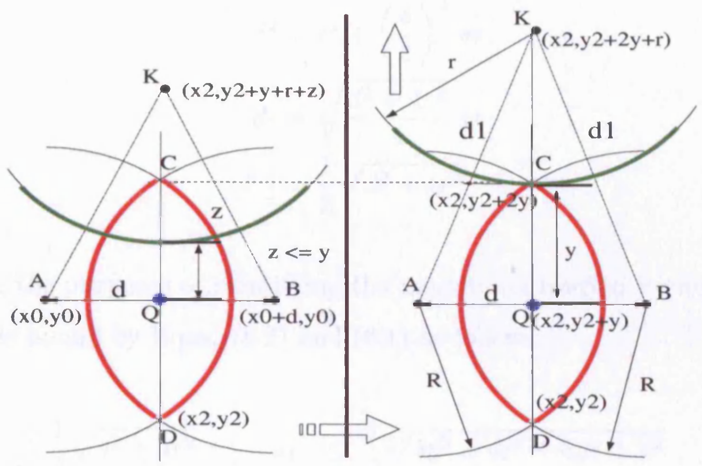


Figure 6.2: 3-cell overlap model.

We introduce a second measure of inter-AP distance d_1, identified as *centre CA adjacency distance* between an CAs neighbour and the centre CA of an M-neighbourhood; in figure 6.2, d_1 is the distance between CA_K and CA_B as well as CA_K and CA_A. Within a 3-cell overlap model, d is identified as *neighbour CA adjacency distance* and is assumed to be homogeneous for all neighbouring CAs (such as CA_A and CA_B) surrounding the centre CA (CA_K) of the M-neighbourhood.

In this manner d and d_1, track the measure of area overlap between 3 CAs as a function of inter-AP distance in two dimensions, namely $A_{overlap} = f(d, d_1)$; d_1 has an upper bound[11] of:

[10] polygons emerging from triangulation of an area, i.e. triangles

[11] through application of Pythagorean theorem for triangle $\triangle AKQ$.

$$
\begin{aligned}
d_1^2 &= (y+r)^2 + \left(\frac{d}{2}\right)^2 \Leftrightarrow \\
d_1^2 &= y^2 + r^2 + 2yr + \frac{d^2}{4} \Leftrightarrow \\
\lceil d_1 \rceil &= \frac{1}{2}\sqrt{4y^2 + 4r^2 + 8yr + d^2} \qquad (6.1)
\end{aligned}
$$

To monitor the effect of *centre adjacency distance* between neighbouring CAs and the centre CA of an M-neighbourhood, we maintain a constant cell radius of cell K while we vary its centre location on the y-axis by z, such that $z \leq y$ where y is calculated according to Eqn. (G.4). This implies that the lower bound for $\lfloor d_1 \rfloor$ is:

$$
\begin{aligned}
d_1^2 &= r^2 + \left(\frac{d}{2}\right)^2 \Leftrightarrow \\
d_1 &= \sqrt{\frac{d^2 + 4r^2}{4}} \Leftrightarrow \\
\lfloor d_1 \rfloor &= \frac{1}{2}\sqrt{d^2 + 4r^2} \qquad (6.2)
\end{aligned}
$$

Hence, for the purposes of identifying the measure of overlap between CA_A, CA_B and CA_K, d_1 is bound by Eqns. (6.2) and (6.1) as follows:

$$\frac{1}{2}\sqrt{d^2 + 4r^2} \;\leq\; d_1 \;\leq\; \frac{1}{2}\sqrt{4y^2 + 4r^2 + 8yr + d^2} \qquad (6.3)$$

3-cell overlap model

The total area of overlap between three coverage areas of *different* radii as the sum of 3 circular segments and a straight edge triangle. By means of trigonometric manipulations presented in Annex G.1 the area of overlap is found to be:

$$A(abc) = A_{CircSegA} + A_{CircSegB} + A_{CircSegC} + A_{\triangle abc} \qquad (6.4)$$

Each circular segment $A_{CircSeg}$ is equal (see Annex G.1) to:

$$A_{CircSegA} = \frac{1}{2}R^2(\theta - \sin\theta) \qquad (6.5)$$

while the triangle area is:

$$A(\triangle abc) = \sqrt{s(s - ab)(s - bc)(s - ac)} \quad (6.6)$$

where s is the semi-perimeter of the triangle. The shape of this area is shown in figure 6.3.

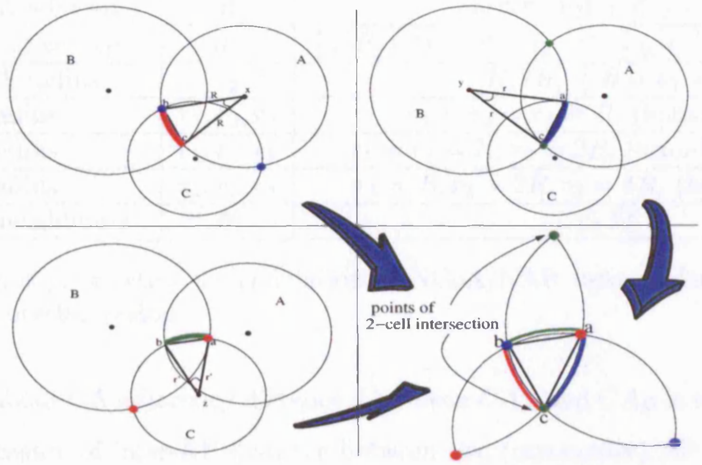


Figure 6.3: Calculating 3-cell overlap as circular segment contribution onto circular triangle

Measure of NCoA/NAR miss probability over 3-cell overlap

To assess the significance of NCoA/NAR miss probability for FMIPv6 we employ the above overlap model into a set of Monte Carlo simulations through Matlab, as a function of three input parameters: (i) neighbour CA adjacency, (ii) centre CA adjacency and (iii) CA radius (centre and neighbour). The bounds of the variables are specified in Table 6.3.

Ransom [356] suggests that to offer significant wireless link service quality, the overlap of coverage areas between adjacent micro-cells, must be considerably large (upto 50%). Wiggard [357] refines such bounds by showing through measurements that the measure of coverage ovelap varies between 10% and 39% the area of both overlapping cells for medium vehicular speeds of mobile hosts. To ensure valid simulation bounds, our overlap simulations are conditioned by the requirement that the measure of coverage overlap varies between 10-40 % of the coverage area of overlapping cells.

We then translate the measure of CA overlap as the respective density function of the NCoA/NAR miss probability as a function of inter-AP distance and cell radius.

The former density function represents the NCoA/NAR miss probability only within a *single* (multi-AP) overlap region of a single CA. To this end, we extend the simulation to consolidate the sum of density functions of NCoA/NAR miss probability for *multiple* (multi-AP) overlap regions of a single CA, as a function of AP neighbours ($f(N)$); we consider typical M-neighbourhood cell configuration with $6 \leq N \leq 10$ AP neighbours, for a constant average neighbour CA adjacency distance $d = 0.7667$.

Type	Variable	value range
neighbour CA adjacency	d	$max(r_1, r_2) \leq d \leq r_1 + r_2$
centre CA adjacency	d_1	$\frac{1}{2}\sqrt{d^2 + 4r^2} \leq d_1 \leq \frac{1}{2}\sqrt{4y^2 + 4r^2 + 8yr + d^2}$
centre CA radius	r_3	$[R, 7R] \mid R = r_1 = r_2$
CA radius	r_1, r_2, r_3	$r_1 = r_2 = r_3 = R$, (homogeneous)
CA radius	r_1, r_2, r_3	$r_1 = r_2 = R$, $r_3 = 2R$, (semi-homogeneous)
CA radius	r_1, r_2, r_3	$r_1 = R, r_2 = 2R, r_3 = 4R$, (heterogeneous)
No. of CA neighbours	N	[6,10]

Table 6.3: Input parameters for simulation of NCoA/NAR miss probability within a variable 3-cell overlap region

The neighbour CA adjacency distance d between CA_A and CA_B is expressed as the normalised measure of inter-AP distance between two (non-centre) AP neighbours (of the M-neighbourhood), over their maximum[12] inter-AP distance (i.e. $d_{max} = r1 + r2$); this is identified as the *neighbour CA adjacency distance ratio* and denoted as $\rho = d/d_{max}$. It follows that, for an increasing ρ the measure of overlap area is expected to decrease; the simulation monitors the behaviour of this decrease.

In a similar fashion, the centre CA adjacency distance d_1 between CA_A and CA_K or CA_B and CA_K is expressed as the normalised measure of inter-AP distance between the centre AP and one of its AP neighbours within the M-neighbourhood, over their maximum inter-AP distance (i.e. $d_{1max} = \frac{1}{2}\sqrt{4y^2 + 4r^2 + 8yr + d^2}$); this is identified as *centre CA adjacency distance ratio* and denoted as $\rho_1 = d_1/d_{1max}$. The expected value of ρ_1 varies in a similar manner to the one of ρ.

For both neighbour as well as centre CA adjacency distance parameters, it is implied that d and d_1 maintain a lower bound, sufficient to preclude any adjacency interference.

Simulation Results

Figure 6.4(a) presents the measure of area overlap as a function of the centre and neighbour CA adjacency distance ratio (ρ_1, ρ), while all 3 intersecting cells maintain the *same* transmission range. It can be seen that for small neighbour (ρ) or centre (ρ_1) CA adjacency distance ratios, the measure of area overlap between 3 cells experiences peak values. It appears that the NCoA/NAR miss probability for a single overlap region

[12]that is, the distance of zero CA overlap.

within a single CA, varies between 2 and 16% for overlapping cells of the same radius, depending on the measure of inter-AP distance.

More importantly, the measure of cell overlap is more sensitive to the centre CA adjacency distance (d_1) than the neighbour CA adjacency distance (d); for instance, for an 80% increase in d, the measure of area overlap between 3 cells reduces proportionaly by half simply by means of a 10% decrease in d_1. This is also reflected in the respective density function of the NCoA/NAR miss probability shown in figure 6.4(b).

The above implies that the positioning of the centre CA with respect to their neighbours within an M-neighbourhood, is significantly more critical in generating larger overlap regions (per 3 cells) and thus a larger NCoA/NAR miss probability, than the positioning of neighbouring CAs with respect to each other. The latter is important in identifying the non-critical parameter (i.e. ρ) for which we choose a constant average value, as we proceed to observe the effects of CA radius onto area overlap and thus onto NCoA/NAR probability.

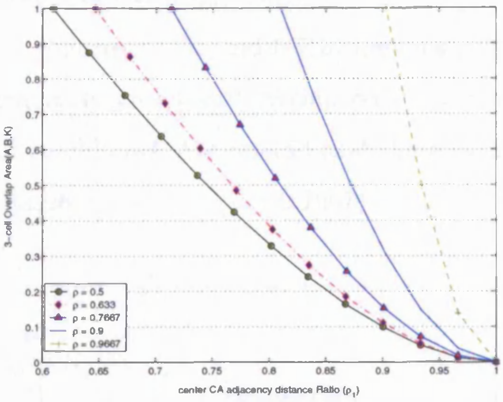
Figure 6.4(c) presents the respective measure of area in a single overlap region within a CA, as a function of its centre CA adjacency distance and the transmission range of its respective AP. It can be seen that the measure of overlap reduces with the *same rate* as the one of figure 6.4(c) only if the radius of the centre CA is significantly *larger* than the one of its AP neighbours.

In practice, such effect dilutes the NCoA/NAR miss probability since a single region of area overlap is accounted over a much larger CA area[13]; the manner by which the NCoA/NAR miss probability is diluted by a significantly larger radius at the centre CA can be seen by comparing figures 6.4(d) and 6.4(b). It appears that a 7-fold increase on the radius of the centre CA can reduce the NCoA/NAR miss probability down to less than 1%. However, such variations in the transmission range of CAs are *not* typical within wireless networks, or last-hop WLAN infrastructures, in particular.

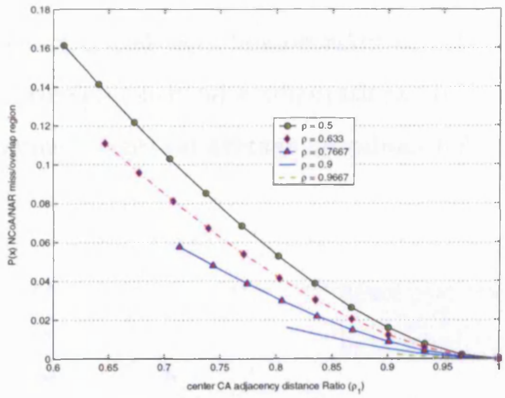
It may be observed from figures 6.4(a) and 6.4(c) that the effect of centre CA adjacency distance (d_1) remains more significant than the transmission range of the individual AP; reducing (ρ), increases faster the area overlap between 3 overlapping cells as opposed to increasing the AP transmission range by the same amount. This becomes obvious from figure 6.4(c), where it requires an increase of the transmission range by a factor of 7 to achieve the same[14] overlap region.

[13]this is because R increases significantly, bringing a proportional increase to the total CA area $A(R) = \pi R^2$

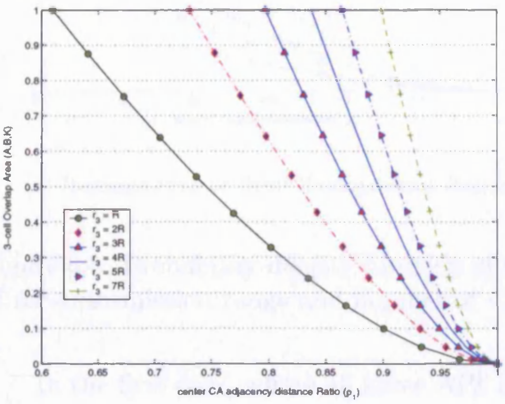
[14]since the normalised centre CA adjacency ratio is achieved over a larger inter-AP distance



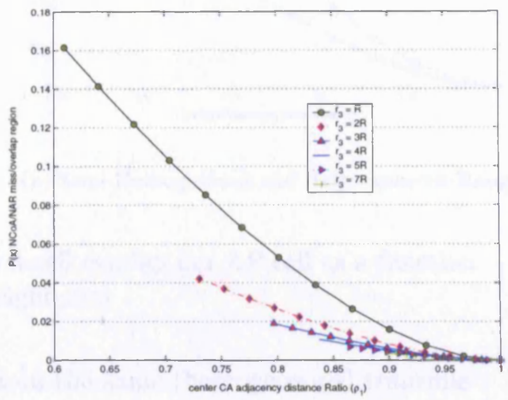
(a) Measure of 3-cell area overlap as a function of $f(\rho_1, \rho)$



(b) Probability Density F(x)



(c) Measure of 3-cell area overlap as a function of $f(\rho_1, r_3)$



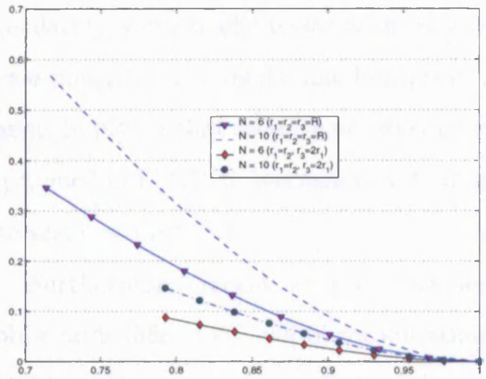
(d) Probability Density F(x)

Figure 6.4: Measure of 3-cell overlap region and its respective probability density function

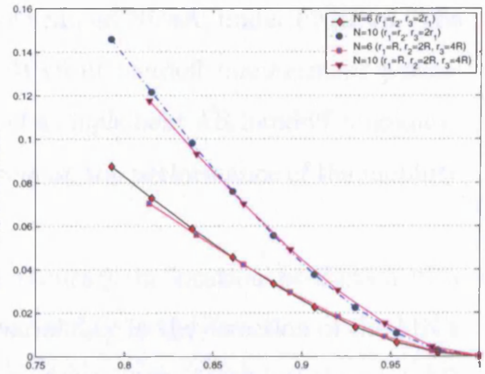
From the above we may conclude that a significant increase in the area of the overlap region is achieved more between CAs with homogeneous transmission range than semi-homogeneous or heterogeneous transmission ranges. The measure of increase in the overlap region tracks subsequently the probability of occurrence of NCoA/NAR miss probability by the respective adjacency distance ratio.

We proceed to explore further this derivation by identifying 2 new simulation scenarios contrasted against homogeneous radii for CAs participating in a 3-cell overlap: (i) semi-homogeneous radii, (ii) heterogeneous radii. We further consolidate our conclusions on the behaviour of NCoA/NAR miss probability, by evaluating its total measure over *multiple* regions of (3-cell) area overlap within a single CA as a function of a varying number of CA neighbours.

Figures 6.5(a) and 6.5(b) demonstrate the effect of cell radii homogeneity or heterogeneity on a 3-cell overlap region for typical M-neighbourhood configurations of N AP neighbours. We remind that these figures assume a constant average neighbour CA adjacency distance ($\rho = 0.7667$).



(a) Homogeneous or Semi-Homogeneous Range



(b) Semi-Homogeneous and Heterogeneous Range

Figure 6.5: Probability density function of total 3-cell overlap per AP cell as a function of AP-transmission range and number of CA neighbours

In the first case, where all three APs maintain the same (homogeneous) transmission range, we observe that for a set of six AP neighbours ($N = 6$) the area in the respective overlap region for all neighbours, yields a NCoA/NAR miss probability of about 0.35. For an increasing number of AP neighbours ($N = 10$) with the same radius, the probability of forwarding the traffic of the MN towards an incorrect NCoA under FMIPv6 increases by about 0.22.

In the second case, where the surrounding CAs reduce their Tx range (semi-homogeneous) by half (i.e. $r_3 = 2r_1$), we observe that the NCoA/NAR miss probability drops (by about 0.26) to 0.095. For an increasing number of neighbouring APs ($N = 10$) around the centre CA (CA_K) of the M-neighbourhood, we observe an increase of the NCoA/NAR miss probability by about 0.06. The former implies that in a semi-homogeneous M-neighbourhood, where the inter-AP distance between AP, *on the edge* of the M-neighbourhood remains constant, while their transmission range is reduced by half, the probability of an NCoA/NAR miss is significantly smaller compared to the case of homogeneous transmission range AP neighbours, but nevertheless non-negligible.

In the third case, where both the surrounding as well as central APs maintain a different transmission range (i.e. $r_1 = R, r_2 = 2R, r_3 = 4R$), we observe only a marginal decrease in the NCoA/NAR miss probability. This is the case for both a small number ($N = 6$), as well as an increasing number ($N = 10$) of AP neighbours.

All three cases of transmission range heterogeneity amongst three neighbouring APs, suggest that the measure of overlap area introduces in the worst case, namely the one of maximum area overlap between three circular cells, a significant probability of encapsulating MN's traffic towards an *incorrectly* determined NCoA, under FMIPv6. The above constitutes a significant limitation for FMIPv6 in handoff management performance, implying that reliance on *exact* prediction of a *single* next AR handoff candidate, as pursued in FMIPv6, has significant adverse effects on the performance of the mobility management protocol.

Furthermore, recent work on the degree of accuracy in location prediction of a mobile node [358, 359, 360] in combination with variability in the direction of the MN's mobility path, suggest strongly that determination of the exact identity of the next AR candidate remains highly inaccurate or at best computationally intensive to achieve. Hence, strong reliance on predictive functions, on the part of the mobility management protocol such as FMIPv6, suggest an equally strong probability of adverse performance in sustaining the seamlessness of MN's on-going interactive real-time flows.

On the contrary, PMIPv6 eliminates any possibility of NCoA/NAR miss probability; in its default (pessimistic) mode, PMIPv6 *does not* predict the exact identity of the next (single) AR handoff candidate. Instead it identifies and communicates with *all* members of the M-R neighbourhood, on the basis that the typical size of an M-neighbourhood is generally small (i.e. $N \leq 10$).

The former, in the light of non-determinism in the direction of MN's path, coupled with varying propagation characteristics of wireless links, allows to identify neighbours and establish IP-roaming state *concretely*[15] for the entire M-neighbourhood with no need for complex, resource intensive AP detection methods on the part of the link-layer as required in FMIPv6.

The above flexibility under PMIPv6 is afforded at the cost of some *additional* L3-signalling on the part of the wire-line network, with state established between the MN and *all* AR neighbours, in contrast to the optimal case of signalling between the MN and the ultimate AR handoff recipient. However, as we show in the following section even under such signalling regime, the measure of signalling cost under PMIPv6 is

[15]state that is verified and can be used with immediate effect

significantly less than that of FMIPv6.

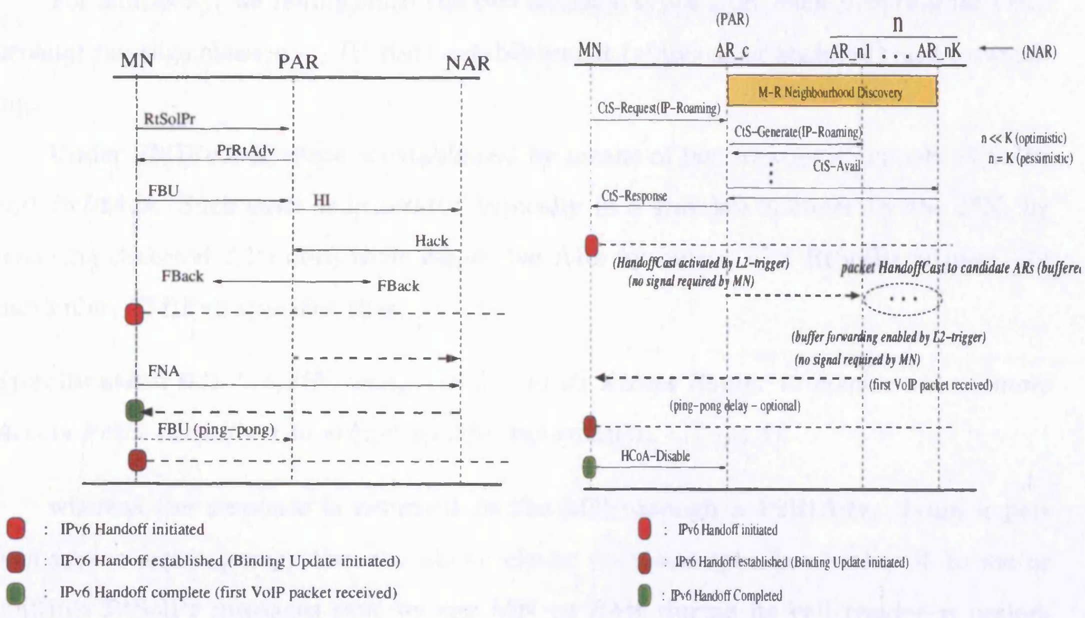
It is emphasised that most of signalling under FMIPv6 is pushed to the MN in an effort to minimise signalling requirements on the part of the AR. This, however, introduces the cost of access contention on the wireless link, as well as reductions in good-put as a result of increased controlled signalling on the part of the MN.

By contrast PMIPv6, in the light of seamlessness, considers imperative that the cost of access contention is minimised; for this purpose, PMIPv6 signalling is pursued over the wireline network (i.e PoAs), with limited signalling requirements on the part of the MN.

6.2.3 Signalling Efficiency

This section presents a comparative analysis of signalling efficiency between FMIPv6 and the proposed PMIPv6 architecture.

We first identify the signalling cost incurred by FMIPv6 and PMIPv6, by describing algebraically the number of signals required by each proposal, from two *perspectives*: (i) the one of the AR, (ii) that of MN. Figures 6.6(b) and 6.6(a) provide a collective view of the signalling incurred by FMIPv6 and PMIPv6. With respect to FMIPv6 our analysis focuses mainly on the case of the 'predictive' handoff signalling. Derivations are then applied on individual components of FMIPv6's alternate case of 'reactive' handoffs.



(a) FMIPv6 Signalling (b) PMIPv6 Signalling

Figure 6.6: Core signalling for Proactive and Fast MIPv6

Under PMIPv6, signalling is classified according to three broad classes of messaging, all at the IP layer: (i) M-R Neighbourhood discovery[16], (ii) state establishment (iii) forwarding (HandoffCast). Under PMIPv6 all state (IP-Roaming or other) established remains verified; that is, the MN can use it with immediate effect with no need for additional checks by an AR[17].

On the contrary, FMIPv6 *does not* specify a method for discovering neighbouring ARs/APs; instead, it assumes that such information is readily available. As such, FMIPv6 encompasses two marginally different message classes: (i) state resolution signalling for generation of routing state and addressing state, (ii) forwarding signalling (unicast). We note that in FMIPv6, given that address generation is predominantly stateless, addressing state remains *unverified* until the MN signals its departure from its current PoA.

Given that FMIPv6 does not specify a mechanism for discovery of AP/AR Neighbourhood state, we exclude the M-R Neighbourhood discovery signalling of PMIPv6 from subsequent analysis and assume, *a priori*, both proposals to have available such information where necessary. Furthermore, the analysis is limited to IP-Roaming state, that is, L2/L3 addressing as well as L3 routing information solely for the purposes of the IPv6 handoff; this is because FMIPv6 state establishment signalling is specifically limited to resolution of L3 addressing and routing information.

For simplicity, we homogenize the two message types from each proposal into two broader message classes: (i) IP-state establishment (stateless or statefull) and forwarding.

Under FMIPv6, IP-state is established by means of two messages, namely *RtSolPr* and *PrRtAdv*. Such state is generated typically in a stateless manner by the MN, by resolving detected APs onto their respective ARs by means of a RtSolPr request. In particular, FMIPv6 specifies that:

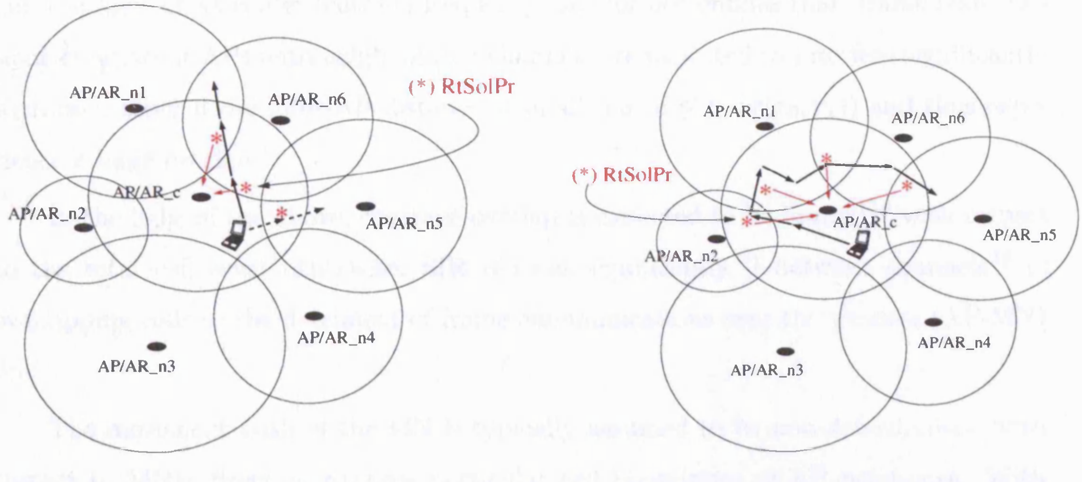
Specification 6.2 *'..a MN sends RtSolPr to its Access Router to resolve one or more Access Point Identifiers to subnet-specific information. ..' (pg.6)*

whereas the response is returned to the MN through a PrRtAdv. From a performance analysis perspective the above clause does *not* specify whether it is one or multiple RtSolPr messages sent by the MN to PAR during its cell residence period. The specification does not, however, exclude the possibility that:

[16] see Section E.2

[17] or other relevant state establishment authority for that purpose.

Specification 6.3 *'..The MN can also send RtSolPr more than once during its attachment to PAR..' (pg.8)*



(a) Suburban mobility pattern (b) Urban mobility pattern

Figure 6.7: Multiple PrSolRt messages generated by FMIPv6 in the hypothetical case of simultaneous AP availability detection by the MN, while in ongoing packet communications with its peers.

It appears, however, that the sending of *multiple* RtSolPr messages under FMIPv6 is, in fact, not a possibility but a matter of necessity for AP-identifiers that may be discovered. We show that an MN is guarranteed to require more than one RtSolPr messages during its cell residence at the current PoA depending on the terrain environment. A minimum of one RtSolPr message may be expected in the case of suburban wireless network envirnoments, as shown in figure 6.7(a). For urban wireless network deployment, the number of RtSolPr messages is expected to vary on average up to $N/2$, where N is the number of AP neighbours with respect to MN's current PoA; this is justified by the fact that in urban wireless network, MN's mobility path is typically not a straight line across AP boundaries [361].

To this end, we first assert that the MN is expected to encounter overlap regions between its current AP and multiple AP neighbours crossing MN's movement path. Typically, however, there exists *no* unique location within the range of the associated AP (in any direction), where the MN can detect immediatelly *all* AP neighbours omni-directionally; this is because, neighbouring APs in any direction, are installed to support a limited measure of overlap in proportion to the cell's total area. Such overlap aims to sustain continuity in wireless coverage, while preventing a low inter-AP signal to interference ratio (SIR) as a result of co-channel interference [328].

With respect to WLAN networks in particular, we recall from Section D.6.1 of Chapter 3, that the RF design of a 802.11 host spreads each channel 5 MHz to the left and the right of its center (carrier) frequency; the former implies that, frame transmissions of adjacent APs with neighbouring channels are expected to interfere significantly with each other if the inter-AP distance is small (i.e. $d \leq max(r_0, r_1)$) and thus experience a large overlap.

In the light of the above, coverage overlap is expected to be limited[18] with respect to the total cell area; otherwise, SIR reduces significantly [] between channels[19] of overlapping cells to the detriment of frame communications over the wireless (AP-MN) link.

The movement path of the MN is typically assumed to be non-deterministic with respect to MN's direction towards particular cell boundaries of AP neigbours. With this in mind, and assuming limited overlap at cell boundaries, the mobility path of the MN is *guaranteed* to incur under FMIPv6 *multiple* RtSolPr/PrRtAdv message pairs between two conseqtutive handoffs. Figure 6.7(b) demonstrates a typical movement path of an MN within an urban environment. If c_s is defined to be the number of RtSolPr/PrRtAdv message handshakes then it can be seen that its magnitude is bound by:

$$1 \leq c_{sFMIPv6} \leq \frac{N}{2} \tag{6.7}$$

for the average case of MN's movement path trajectory, whereby on average $N/2$ neighbouring APs are detected or

$$1 \leq c_{sFMIPv6} \leq N - 1 \tag{6.8}$$

as a conservative worst-case scenario of MN's mobility pattern where by the MN detects all AP neighbours except for one: its previous PoA; we identify this as *postman movement*. The aforementioned measure of state establishment signalling describes the signalling cost from perspective of the MN; the respective cost from the perspective of the AR becomes:

[18]such limits (inter-AP distance) have been defined in Table 6.3

[19]We remind that channel allocation in WLANs is deregulated with only a 3-channel reuse pattern; thus the only real safeguard to interference in dense WLAN environment is limited overlap

$$m \leq c_{sFMIPv6} \leq m\frac{N}{2} \tag{6.9}$$

while for postman movement effects on the mobility pattern of the MN, such cost becomes:

$$m \leq c_{sFMIPv6} \leq m(N-1) \tag{6.10}$$

where m is the number of mobile nodes attaching temporally at that PoA while in transit. The design consideration of such signalling regime under FMIPv6, is intuitive: attempt to minimise AP resolution signalling on the MN by employing it *on-demand*, that is, when APs are detected and thus, signifying potential AR handoff candidates.

From the perspective of the MN, PMIPv6 incurs a single CtS-Request/ CtS-Response for IP-Roaming state per MN for the entire set of AR members of its R-neighbourhood. However, from the perspective of the AR, it incurs the constant cost of:

$$c_{sPMIPv6} = m + N \tag{6.11}$$

With respect to forwarding and from the perspective of the MN, FMIPv6 incurs a constant cost of one handshake for forwarding initiation. From AR's perspective the constant cost rises to $2m$, since one handshake is expended in the FBU/FBack signals and one for the HI/Hack signals, for m MNs. Furthermore, forwarding terminates with an FNA which accounts for 0.5 handshakes that conclude the handoff process. This reshapes the signalling cost of FMIPv6 forwarding at the AR to $2.5m$ handshakes.

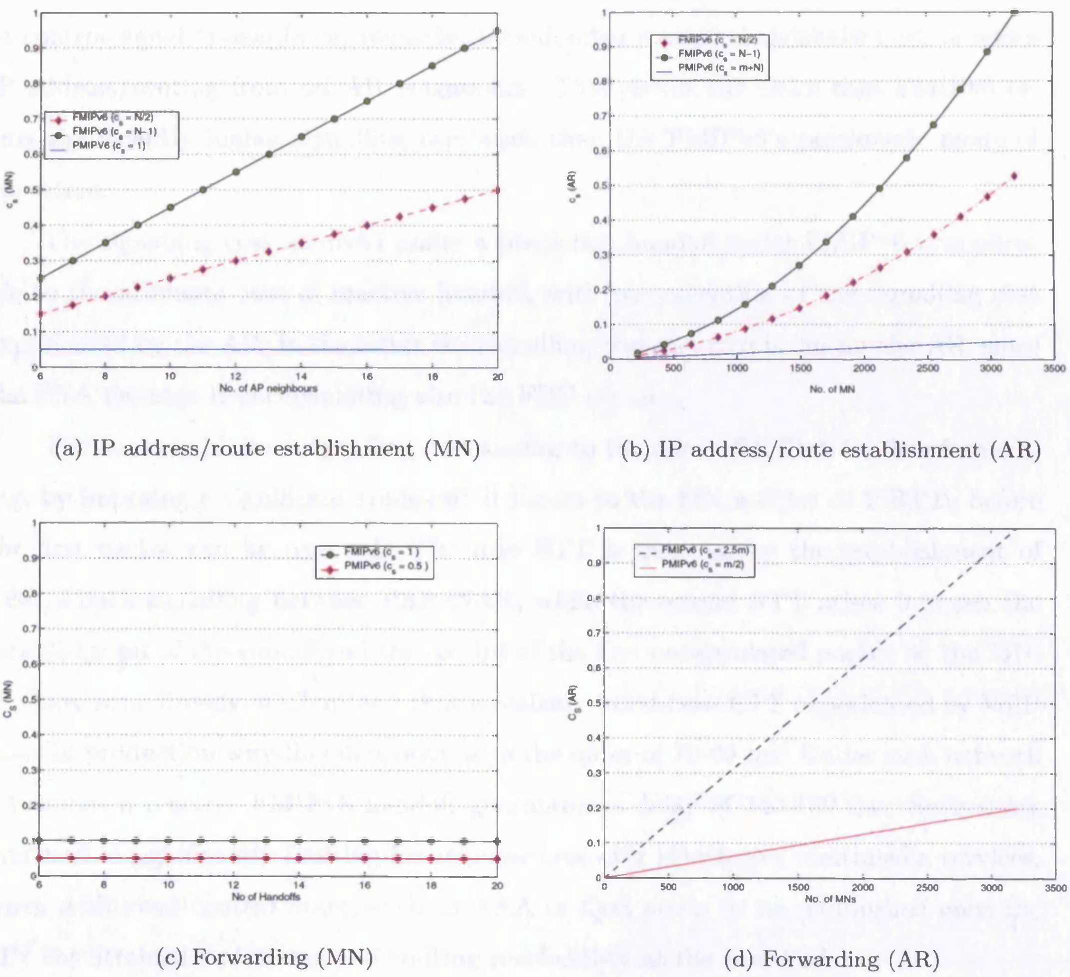
PMIPv6 on the contrary expends only half a signalling handshake for forwarding termination, since the activation/deactivation of the forwarding function is effected through link-layer triggers. From the perspective of AR the forwarding signalling cost, under PMIPv6 rises to a mere $m/2$.

It may also be reminded that the signalling cost incurred by PMIPv6 is fixed since the establishment IP-Roaming state (e.g new CoA) is guaranteed to be verified. This is not the case for FMIPv6 under stateless address auto-configuration conditions. The signalling cost for both PMIPv6 and FMIPv6 proposals is summarised in Table 6.4. It may also be seen that both FMIPv6 and PMIPv6 attempt to preserve seamlessness at the cost of increased control signalling (handshakes) in comparison to Mobile IPv6.

Signalling Cost	MIPv6		FMIPv6		PMIPv6	
Perspective	MN	AR	MN	AR	MN	AR
IP Addr/Route Establishment	1	m	$1 \leq c_{sFMIPv6} \leq \frac{N}{2}$	$m \leq c_{sFMIPv6} \leq m\frac{N}{2}$	1	m+N
Forwarding	0	0	1	$2.5m$	0.5	$m/2$

Table 6.4: Signalling Efficiency between FMIPv6 and PMIPv6

Figure 6.8(a) and 6.8(b) present the signalling cost of IP address/routing state establishment from the perspective of MN and AR. Subsequent figures 6.8(c) and 6.8(d) present the signalling cost of forwarding between FMIP and PMIPv6 of MN and AR respectively.



(a) IP address/route establishment (MN) (b) IP address/route establishment (AR)

(c) Forwarding (MN) (d) Forwarding (AR)

Figure 6.8: Signalling cost for IP address/route establishment and forwarding

It can be seen from figure 6.8 that PMIPv6 incurs significantly less signalling in a consistent manner for both state establishment and forwarding message classes. Exception to this, is the case of forwarding overheads experienced from the MN (figure 6.8(c)) where the constant signalling cost increases from one to two signals. What's more, we can observe from figure 6.8(b) that from AR's perspective, the *on-demand* form of AP-AR resolution in FMIPv6 increases its respective signalling cost over the *wireless link*, for a growing number of bypassing MNs, in a *multiplicative* manner; on the contrary, in PMIPv6 the respective signalling increases in an *additive* manner, since each MN contributes a single CtS-Request/CtS-Response for the entire M-neighbourhood.

The result of such multiplicative increase under FMIPv6 has also a negative impact on the contention of the wireless channel, since the number of medium access requests for L3 control signals increase up to $mN/2$, for m MNs and N AR neighbours. Under PMIPv6, contention from control signalling grows linearly with a peak of m control-signal transmission requests, by enforcing a single handshake that provides IP address/routing from *all* AR neighbours. This proves our claim that FMIPv6 incurs significantly higher signalling overheads than the PMIPv6's pessimistic mode of operation.

The signalling cost incurred under a predictive handoff under FMIPv6 is applicable to the alternate case of reactive handoff, with the exception of the signalling cost experienced by the AR; in the latter the signalling cost incurred is $2m$ for the AR, since the FNA message is encapsulating also the FBU signal.

The above exhibits a signalling cost similar to the one of PMIPv6 for flow forwarding, by imposing a significant trade-off: it incurs to the MN a delay of 2 RTTs before the first packet can be received. The first RTT is incurred by the establishment of FBU/FBack signalling between PAR/NAR, while the second RTT arises between the establishment of the tunnel and the receipt of the first encapsulated packet by the MN. We have seen already in Chapter 3 that a realistic worst-case RTT experienced by VoIP flows in production wire-line networks, is in the order of 70-80 ms. Under such network conditions a reactive FMIPv6 handoff guarantees a delay of 140-160 ms. Such delay overhead is significantly limiting for seamlessness over interactive multimedia services, when additional context state, such as AAA or QoS needs to be established once the MN has attained addressing and routing reachability at the new PoA.

6.2.4 IP-Roaming state establishment timing

The sending of an RtSolPr message, under FMIPv6, is highly dependent on the availability of AP-identifier information through some hypothetical 'link-specific' (trigger) event available from MN's link-layer.

It is intuitive that, the rate of arrival of detected AP-identifier information at the MN, is further dependent on MN's crossing within the transmission range of an AP neighbour, as well as the number of AP neighbours within its M-neighbourhood. This is typically manifested as the amount of cell overlap between the CA of the AP currently accommodating the MN and its AP neighbours, crossing MN's mobility path. Given the non-determinism in the mobility pattern of an MN and for the purposes of analysis, we have identified three distinct cases of MN movement with respect to AP discovery, after MN's last IP handoff: (i) straight (suburban) movement (see fig. 6.7(a)), incurring one to two detected APs, (ii) non-straight (urban) movement (see fig. 6.7(b)) incurring on average $N/2$ AP detections, (iii) heavy (urban) movement (postman movement) incurring an average of $N-1$ AP detections.

Under FMIPv6, resolution of AP-identifier information is crucial to the determination of MN's NCoA address for the purposes of subsequent tunnel establishment (FBU transmission) in advance of MN's next IPv6 handoff. It implies that AP-identifier information must be resolved *regularly*, that is through *multiple* RtSolPr messages; this is essential to ensure that the MN receives in a timely manner the resolved AR information to determine its NCoA before tunnel establishment (FBU) is signalled to AR_c. It is noted that a single RtSolPr for *all* detected AP-identifier information within MN's residence time within a cell is precluded, since the MN *cannot* predict with certainty the timing of its next IP handoff, due to continuous variations in movement and propagation characteristics.

In the light of the above and with respect to the timing of an RtSolPr message the FMIPv6 draft specifies that:

Specification 6.4 *'.. The MN may send RtSolPr at any convenient time, for instance as a response to some link-specific event (a 'trigger') or simply after performing router discovery. ..' (pg. 7)*

The above specification does not make clear the default timing of an RtSolPr send to AR_c. To aid elaboration, it is assumed that an RtSolPr is generated *immediately* after some hypothetical link-specific method has signalled the detection of one or more APs in a single scan.

We study the effect of such timing in the respective control signalling over the performance of random access over a single channel multi-access control mechanism. The measures of performance sufficient for the purpose of this analysis are limited to the probability of successful transmission of an MN as well as the probability of collision between transmitting MN associated with the same AP. We compare such measure with the respective probability of collision arising under PMIPv6 as a result of a single signalling exchange during MN's IP-Roaming state establishment.

Let m be the number of MNs associated with a single AP, n is the number of slots comprising the available residence time at an AP for all MNs. Let j be the slot selected randomly by an MN for its transmission. For the perspective of collision probability, let k be the number of MNs attempting to transmit simultaneously over the j_{th} slot. It is assumed that each MN performs exactly one transmission (that of the control signal) which consumes exactly one slot accounted as unit time.

For a fixed number of m MNs we note each j_m of the available n slots with sample slot probability $(1/n)^m$. For all slots $j_1, \ldots, j_{m-1}$ remain $n - j$ possible random slots. Hence the probability that an MN transmits on the j_{th} slot becomes equal to:

$$P_s(x = j) = \left(\frac{n-j}{n}\right)^{m-1} \left(\frac{1}{n}\right) \tag{6.12}$$

The respective collision probability between two or more MNs transmitting at the same time, may be found by considering the equivalent problem of putting m balls into j out of the n available cells [213]; in this manner we find the probability that k or more balls (MN's transmitting each a single message) occupy cell (slot) j_i at the same time, where $k = \{2, \ldots, m\}$; k balls can be chosen in $\binom{m}{k}$ ways, and placed in j cells, while the remaining $(m - k)$ balls may be placed in the remaining $n - j$ cells in $(n - 1)^{m-k}$ ways. For instance for $k = 2$ the collision probability is

$$\begin{aligned} P_s(k = 2) &= \binom{m}{2} \left(\frac{1}{n}\right)^m (n - j)^{m-2} \\ &= \binom{m}{2} \left(\frac{1}{n}\right)^2 \left(\frac{n-j}{n}\right)^{m-2} \end{aligned} \tag{6.13}$$

for $k = (k - 1)$ the respective collision probability is

$$P_s(k = (k - 1)) = \binom{m}{k-1} \left(\frac{1}{n}\right)^{k-1} \left(\frac{n-j}{n}\right)^{m-k-1} \tag{6.14}$$

From 6.13 and 6.14, it follows that the probability that two or more MNs collide during an RtSolPr transmission becomes the sum of all individual probabilities as follows:

$$P_c(m \geq k \geq 2) = \sum_{k=2}^{m} \binom{m}{k} \left(\frac{1}{n}\right)^k \left(\frac{n-j}{n}\right)^{m-k} \tag{6.15}$$

Equations 6.12 and 6.15 describe the probability of success and collision, respectively, for m MNs each transmitting a single (control) message at a random time slot j with cell residence period n. As such these are applicable in the case of PMIPv6 control signalling effected for IP-Roaming state establishment. Under PMIPv6, these probabilities remain the same for *any* mobility pattern since it incurs contently a single control signal irrespective of the path trajectory followed by the MN during its cell residence period.

From Eqn. 6.12 and 6.15 we can derive the respective probabilities for FMIPv6 for urban and postman movement mobility patterns. To do this we first identify each MN continues to transmit one message at a time. However, the number of transmitted messages (balls) to be placed in n slots (cells) increases from m to $(m * N/2)$ for an urban mobility pattern and $[m(N-1)]$ for postman movement under FMIPv6. Thus, in the case of urban mobility pattern the probability of successful transmission at the j_{th} slot is:

$$P_s(x = j) = \left(\frac{n-j}{n}\right)^{m\frac{N}{2}-1} \left(\frac{1}{n}\right) \tag{6.16}$$

while the probability of collision becomes:

$$P_c(m \geq k \geq 2) = \sum_{k=2}^{m} \binom{m\frac{N}{2}}{k} \left(\frac{1}{n}\right)^k \left(\frac{n-j}{n}\right)^{m\frac{N}{2}-k} \tag{6.17}$$

For postman movement the respective success and collision probabilities are:

$$P_s(x = j) = \left(\frac{n-j}{n}\right)^{m(N-1)-1} \left(\frac{1}{n}\right) \tag{6.18}$$

while the probability of collision becomes:

$$P_c(m \geq k \geq 2) = \sum_{k=2}^{m} \binom{m(N-1)}{k} \left(\frac{1}{n}\right)^k \left(\frac{n-j}{n}\right)^{m(N-1)-k} \tag{6.19}$$

Equation 6.15 represents the collision probability of a message transmitted over the wireless link over PMIPv6, while 6.17 identifies the average collision probability for signalling over FMIPv6. The exponents reveal instantly that the collision probability of control signalling over FMIPv6 is significantly larger than the one of PMIPv6 over the last hop wireless link.

Furthermore, completion of router discovery from the perspective of RtSolPr transmission timing, implies receipt of a router advertisement message according to [103]. Clearly, in this case that the sending of an RtSolPr message under FMIPv6 depends on receipt of Router Advertisement messages; from the perspective of timing, however, such dependency does not provide any particular performance benefits; it effectively allows the sending of an RtSolPr at any time during MN's cell residence period at the new PoA.

PMIPv6 does not impose such a dependency either by means of existence of link-specific event or by router discovery; the MN refreshes its IP-Roaming state with a single control signal after MN's handoff has completed; namely, when bindings have been updated with MN's peers. In fact, PMIPv6 *eliminates* the need for RtSol/RtAdv messages for mobility management purposes altogether.

6.2.5 AP Neighbour Identification Efficiency

Prior to resolution of AP-identifier information, discovery of 'available' AP neighbours must be performed. To this end, FMIPv6 specifies that:

Specification 6.5 *'..the expectation is that prior to sending RtSolPr, the MN has discovered the available APs by link-specific methods. ..' (pg.7)*

What is also important, however, is that the process of AP discovery does not impact adversely any on-going, interactive, real-time communications. The capability of combining AP discovery while sustaining packet communications relies fundamentally on the type of wireless link employed by the respective wireless technology at hand. Two abstract types of wireless link are typically available among wireless technologies: (i) half-duplex, (ii) full duplex. We show, that AP discovery over *half-duplex* wireless links impacts negatively the performance of on-going interactive real-time services in production wireless network infrastructures.

A half-duplex wireless link employs a single frequency for both transmission and frame reception. On the contrary, full-duplex transceivers in wireless technologies employ typically two frequencies, one for the up-link and one for the down-link. Full-duplex is inherently expensive to implement and requires typically, frequency planning [362]; the latter imposes an upper limit on the amount of available channels with acceptable frequency separation for both up-link and down-link communication.

Typical wireless technologies with full-duplex transceiver capabilities are cellular networks [328]. These are frequency-regulated wireless infrastructures, whereby channel pairs allocated to wireless nodes are frequency-planned, on a per-cell basis. In cellular networks, pilot signal reception from multiple BSs allow immediate resolution of the next PoA of the MN; this is achieved by means of dedicated control channels that do *not* interfere with the operation of the terminal's allocated data channels; as a result, the exact AP may be identified in parallel with packet communications effected over MN's data path and potentially resolved, under FMIPv6, with a single RtSolPr/PrRtAdv handshake.

On the contrary, wireless technologies operating in the 2.4 or 5 GHz ISM band, like 802.11b/g/a WLANs, are exclusively half-duplex and thus cannot support reception while transmitting.

Such a constraint limits significantly the ability of the MN pursue AP discovery[20] while engaging in active communications with its peers over the shared channel. In particular, discovery of 'available' APs *blocks* the transmission of 802.11 frames (i.e IPv6 packets) on the wireless link [354] for the *entire* scanning period of *all* available channels. This pertains to the fact the MN needs to: (i) enter into a monitor mode during which packet transmission is precluded (ii) discover APs typically by scanning through channels that are different than the one associated with the current AP, (iii) a scan can discover an AP only if propagation conditions with that AP permit to do so.

Techniques, such as *interleaving* AP discovery with frame Tx/Rx while promising appear to suffer [354, 336] from buffer exhaustion on the AP and subsequent attendant packet loss. Derivatives of interleaving techniques such as Syncscan [335] assume perfect synchronisation of AP clocks on a millisecond basis. Furthermore, the effectiveness of such method depends on the MN scanning frequency per channel as well as the dwell period over each channel. Ramani [335] suggests a nominal scanning period of 500ms with a typical wait period of 20-40 ms depending on the hardware implementation. Given that there are 11 available[21] 'channels', a complete scan requires about 5 sec before all channels are probed. Hence, a delay of 220-440ms or 11-22 lost packets[22] is incurred for each complete scan of the channel space, every 5 seconds; we note that the wait period remains the same even if the nominal scanning period reduces by half;

[20] traditionally effected during an L2-handoff
[21] 13 in US (FCC).
[22] assuming a packetisation rate of 20ms

in fact for a shorter scanning period the rate of delay or accompanying loss increases proportionally.

Furthermore, for an AP to be detected, the MN *must* remain, within the overlap area of the relevant AP neighbours during the scan of *all* 11 channels. Even if interleaved there is no guarantee that an full channel scan can provide accurate information on available AP neighbours, since the MN is in motion.

From the above it can be seen that, AP discovery in 802.11 networks, by means of link-specific methods, can incur significant disruption on interactive real-time communications between the MN and its peers. On the contrary, PMIPv6 is independent of the wireless technology, since it does not attempt to scan and resolve L2-specific information such as AP beacon signals. PMIPv6's M-R neighbourhood discovery function is specifically designed to provide a concrete AP/AR neighbourhood mapping with minimal signalling cost[23] with no requirement for complex and disruptive AP detections methods that can affect adversely the performance of MN's on-going interactive multimedia flows.

6.2.6 Tunnel Setup Efficiency

So far we have shown in Section 6.2.1 (Spec.6.1) of the FMIPv6 draft that, the resolution of multiple AP identifiers to AR information does not provide any assurance that some particular AR emerges as the primary candidate for MN's next IPv6 handoff. However, FMIPv6 specifies further that:

Specification 6.6 *'.. With the information provided in the PrRtAdv message, the MN formulates a prospective NCoA and sends an FBU message. The purpose of FBU is to authorise PAR to bind PCoA[24] to NCoA, so that arriving packets can be tunnelled to the new location of the MN. ..' (pg.7)*

Given the availability of multiple subnet-information from resolved AP-identifiers according to Spec.6.1, it follows that the above specification cannot determine with certainty the *correct* prospective NCoA of the MN. This has a cascading effect on determining which is MN's correct new PoA so that: (i) a tunnel can be established at the correct NCoA by means of a subsequent FBU, (ii) packets can be tunnelled to the correct NAR.

[23]1 handshake between AR neighbours, from previous AR information provided by the MN as presented in Section 4.7 and 4.7.1.

[24]Previous CoA

Thus, during the stage of tunnel establishment to a single NAR, FMIPv6 faces significant limitations in identifying the correct new PoA before PAR can setup a tunnel to MN's new NCoA. It is noted that mult-tunnel establishment to MN's unicast candidate NCoAs, is guaranteed to be inefficient, due to multi-tunnel setup overheads per MN and packet replication on paths towards multiple PoA destinations. On the contrary, the PMIPv6 is guaranteed to generate one and only one HandoffCast tunnel based on MN's verified HCoA address; HandoffCast ensures by means of multicast routing that multiple copies of MN's traffic are eliminated towards its new PoA.

State establishment on ARs

With respect to establishment of state, the FMIPv6 draft claims that:

Specification 6.7 .. *The RtSolPr and PrRtAdv messages do not establish any state at the access router..(pg.6)*

The statement is found to be *false* by subsequent specification clause found in the same FMIPv6 specification. We focus on the most important one. In particular, FMIPv6 specifies that:

Specification 6.8 *..If the PAR does not have an entry corresponding to the new access point, it MUST respond indicating that the new access point is unknown..(pg.9)*

The above indicates that AR_c (PAR) *must* maintain some form of state (table entry) providing a mapping between some AP identifier and its respective AR neighbour. Such state, allows to match an AP-identifier received through an RtSolPr message, against *an entry that maps the new AP onto its AR*, implying router-prefix entry (i.e. subnet-specific information).

Clearly, AP-identifier information does *not* typically exist in routing tables. The mapping between an AP-ID and its respective subnet-specific info *must* be stored at the AR (whether PAR or NAR[25]). In this light, an RtSolPr message *does* introduce some state to be stored at the AR. This is similar to state stored on ARs by PMIPv6, in the form of RNV-MNV vector mapping between the AR and the AP constituent of a PoA.

From the above we can conclude that both FMIPv6 and PMIPv6 proposals require state maintenance on AR neighbours for the purposes of MN state establishment.

6.2.7 Tunnel Activation

Typically under FMIPv6 an FBU is sent any time after receiving a PrRtAdv. However, FMIPv6 does not specify the timing of an FBU message sent by the MN to PAR. This

[25] Each NAR is a PAR for its set of attached MNs

is critical to the performance of a handoff mechanism; an FBU establishes immediately both disruption of received traffic at its current PoA and tunnel forwarding of MN's traffic towards its NCoA, according to the respective specification clause:

Specification 6.9 *'..the PAR MUST continue to forward packets to the MN on its current link until the FBU is received. ..' (pg.7)*

In fact, the FMIPv6 specification considers the dispatch of an FBU a trivial event by assuming an ideal link-layer trigger. In particular:

Specification 6.10 *'.. The FBU SHOULD be sent from PAR's link whenever feasible. For instance, an internal link-specific trigger could enable FBU transmission from the previous link..' (pg.6)*

From a performance analysis standpoint the above cannot be accounted without significant delay or packet loss being incurred, after the dispatch of an FBU signal onto MN's on-going interactive flows with its peers. An FBU sent *too early* before the handoff of the MN will cause traffic to be tunnelled to the NCoA configured by the MN at the candidate NAR. During this period: (i) the MN would not receive any traffic (since it is tunnelled to the NCoA) (ii) while the tunnel traffic is buffered at NAR, its delay variance (jitter) begins to increase unacceptably, (iii) buffered packets would miss their play-out scheduling deadline simply by having the MN sending the FBU 500ms earlier than the ultimate handoff of the MN (iv) buffer exhaustion at the NAR will result in guaranteed packet loss. An FBU sent *too late* with respect to the timing of MN's IPv6 handoff, will cause the FBU to be lost. This in turn will cause MN's flow to experience a temporal black hole effect, with packet loss implications until the MN attaches reactively on the new wireless link.

Furthermore, the variability of propagation characteristics in a wireless environment makes extremely difficult to predict the *exact* period for FBU transmission; such prediction measurements require accounting for MN's instantaneous velocity per frame transmission, the buffer capacity on the neighbouring AP/AR handoff candidate, as well as the angle of direction, with respect to the intra-cell residence location at which the MN dispatched the FBU signal. Most of the aforementioned metrics present a prohibitive management cost; hence, they are typically *not* accounted or managed in anyway by the link-layer of most wireless technologies.

We should also note that the assertion by the FMIPv6 specification, positing that there exists a link-layer trigger such that it can predict *when* the link margin will be exhausted and thus, cause an IP-handoff, appears to be unattainable for the purposes of FBU signalling. Non-determinism in the mobility pattern of the MN or propagation effects, can offset with strong probability [28] the prediction of such timing with cascading effects on the timing of FBU signalling.

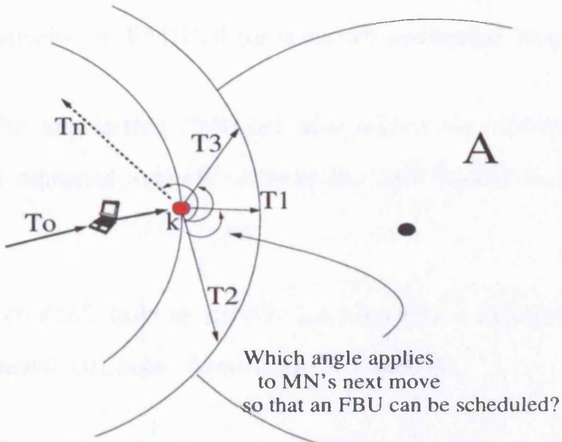


Figure 6.9: Uncertainty on determination of FBU dispatch period before a disruptive SIR that causes the MN to detach from its current CA.

Figure 6.9 presents a potential IPv6 handoff scenario from PoA A to B; while the MN is departing from direction T_o, it encounters a hypothetical critical soft-handoff point whereby it has to schedule an FBU. Even by aid of some ideal link-layer trigger, the MN or its link layer cannot predict which of the directions $\{T_1, T_2, T_3, T_n\}$ the MN will follow; more accurately, which of the respective angles will be effected in MN's next move? Here the ideal link-layer trigger asserted by FMIPv6, fails to provide an accountable answer even if it can monitor MN's velocity on a continuous basis. It is reminded that each of the directions mentioned above are bound to have different buffer requirements, but above all a different impact on the typical delay bound of 200ms of MN's interactive multimedia flows (e.g. VoIP).

The above suggest ultimately that: (i) such an ideal link layer trigger does not exist or at best (ii) any well-defined link-layer trigger is bound to introduce unavoidable delays which for interactive services and above the nominal bound of 200ms, translate to packet loss.

PMIPv6 does not introduce such complexities; the HCoA activation/suspension mechanism can afford to be conservative since traffic can *also* be received at MN's current PoA, while forwarded to MN's HandoffCast (HCoA) address.

For robustness, packet duplication on MN's current PoA, is handled by conditional forwarding over the wireless link, subject to the availability of an `ATTACH` L2-trigger.

6.2.8 NAR Neighbourhood Determination

We have seen from the Spec. (6.1) that in the case of *MN-controlled* handoffs, a solicited PrRtAdv provides the MN with one or more [AP-ID, AR-Info] tuples. A similar specification clause is identified in FMIPv6 for *network-controlled* handoffs, whereby:

Specification 6.11 *..The unsolicited PrRtAdv also allows the network to inform the MN about geographically adjacent subnets without the MN having to explicitly request that information..(pg.7)*

However, the FMIPv6 draft fails to specify (or identify) a mechanism for AR discovery between such adjacent subnets. According to FMIPv6:

Specification 6.12 *..The method by which Access Routers exchange information about their neighbours and thereby allow construction of Proxy Router Advertisements with information about neighbouring subnets is outside the scope of this document..(pg.10)*

On the contrary, PMIPv6 establishes a M-R Neighbourhood discovery [50, 52] as the very core of its mobility management approach. Recent work by Chalmers et al. [147] confirms that the M-R Neighbourhood discovery mechanism devised by the proposed Proactive MIPv6 architecture is the first to set the foundations for adjacency information exchange. We augment, this confirmation by arguing that the proposed M-R Neighbourhood discovery mechanism extends beyond adjacency information exchange towards capability exchange in any context or ontology related to mobility management.

DAD Handling

With respect to duplicate address detection FMIPv6 specifies two potential functions applicable at the AR_n (NAR): (i) AR_n has special knowledge of *all* MN addresses under its subnet (ii) AR_n does not have such special knowledge. The FMIPv6 specification fails to identify which is the default mode of operation and which the optional. To this end, both alternatives are analysed, with particular emphasis on the repercussions of each approach on the complexity of subsequent implementation.

In regards to the first potential mechanism of DAD resolution, the FMIPv6 specifies that:

Specification 6.13 *..the NAR can have a list of all nodes on its subnet, perhaps for access control, and by searching this list, it can confirm whether the MN's address is a duplicate or not..(pg.14)*

While the above may appear to be a convenient solution it puts reliance on a mechanism *external* to the Neighbour Discovery standard that has been traditionally used for DAD resolution. If this is supposed to be the default mode of operation, given that DAD resolution is a mandatory function, irrespective of the low probability of unicast address collision, then an FMIPv6 implementation would require the implementation of a special access control list for DAD purposes *on a per-subnet* basis. Such an assumption appears to be unrealistic, since FMIPv6 implies by design, support of access control, or alternatively, an interface between the FMIPv6 specification and some well-specified access control mechanism. On both accounts FMIPv6 fails to identify which of the two modes it is supporting and by means of what state or signalling. It is clear, nevertheless that in such case the DAD function would incur either the generation of additional state or additional control signalling[26].

With respect to the second potential form of DAD resolution, the FMIPv6 specifies that:

Specification 6.14 *..If such knowledge is not available at the NAR, it may indicate this by not confirming NCoA in the HAck message. The NAR may also indicate this in the NAACK option as a response to the FNA message. In such cases, the MN would have to follow the address configuration procedure according to [6] after attaching to the NAR..(pg.14)*

If this is the default mode of operation for DAD resolution, it is expected to incur a constant delay of 1000ms in the handoff process of the MN at NAR, as shown experimentally in Chapter 3. Even if buffering is allowed at NAR, the MN is guaranteed to experience, by-design, a packet loss of around 800-850ms from the DAD process alone; this is because during the period of standard DAD resolution (i.e. 1000ms) 800ms worth of VoIP packets would have already exceeded their payout schedule deadline[27], even if the NAR can buffer 1000ms worth of packets per MN. In fact, it can be easily deduced that buffering may yield a saving of only 0.15-0.2 of the delay incurred by the DAD process.

[26] to an access control host

[27] or for a high degree of interactivity in VoIP communication with the peer the loss can be as big as 850ms

On the contrary, in both of the above cases, PMIPv6 imposes *no* such delay during the period of MN's handoff. The MN requests the establishment of IP-Roaming state (including the set of tentative sCoAs[28]) at the PAR which in turn delegates the request to its AR neighbours, candidates for MN's handoff. Generation of MN's sCoA (NCoA) address at each candidate PoA is performed by the NAR through a proxy-stateless address auto-configuration, proposed in Section E.6.1; that is, the NAR configures itself the soft CoA (NCoA) address and submits it for standard DAD resolution on-link, by means of a standard Neighbour Solicitation [107]. Because the proxy-stateless address configuration mechanism is effected well in advance of MN's next IPv6 handoff, the standard DAD delay of 1000ms does not affect MN's packets communications *during* its handoff.

6.2.9 Unicast tunnelling vs HandoffCast

PMIPv6 employs HandoffCast for the purposes of traffic forwarding towards its new PoA within its current M/R neighbourhood. The underlying routing mechanism of HandoffCast is IPv6 Multicast, in contrast to FMIPv6's unicast tunnel. HandoffCast does not alter multicast routing in anyway. Furthermore, it remains independent of the underlying multicast routing protocol.

While the benefits from employing HandoffCast for the purposes of flow forwarding towards MN's new PoA have long been elaborated throughout this Chapter, its weakness lies not on performance, but in deployment.

For instance, the speed of prunes and grafts of new AR neighbours with the maintained R-neighbourhood of the MN is typically fairly low, and for the period of state establishment, well in advance of MN's handoff, negligible. In particular, for sparse mode algorithms [185], the signalling delay of prunes/grafts incurs about 0.5 RTTs between the source and the receivers; for shared-trees multicast algorithm [185] this increases to 1 RTT for the graft between the receiver and the Rendezvous Point (RP) including the subsequent data back to the receiver. For worst-case variability in round trip times an AR prune/graft is not expected to exceed the measure of intra-domain round trip delay (in the order of 40 and 80 ms respectively). Given that such signals are effected well in advance of MN's handoff, it is easy to see that such delays are negligible for mobility management purposes.

However, inter-domain multicast deployment is limited by policy considerations of ISPs on cooperative multicast routing . This is due to the inherent nature of the

[28]The term soft CoA used by Proactive MIPv6 is semantically equivalent to NCoA used by FMIPv6.

underlying multicast forwarding tree; the generation of a shared multicast tree relies fundamentally on what is known as a reverse path forwarding (RPF) check.

What appears to be essential for an ISP (e.g. ISP_A), is to effect control on routing prefixes that are to be used for RPF checks; namely, select what are the routes that constitute the forwarding tree connecting the source to one or more interested destinations. If there exists no policy control over an RPF check, then establishment of the multicast forwarding tree (and thus the availability of multicast routing) is largely dependent on other *external* network domain entities (e.g. ISP_B).

Without BGP policy controls, external ISPs (e.g. ISP_B) are attached unwillingly as part of a multicast forwarding tree setup at domain ISP_A; ISP_B acts essentially as an involuntary carrier of traffic from domain ISP_A. Domain ISP_B maintains no control[29] over this, except for enabling or disabling multicast routing; that is to say, *inclusion of an ISP onto a multicast forwarding tree arises as a matter of algorithmic performance rather than provisioning policy.* Such inclusion is enforced even when the ISP_B has no sources or receivers onto that forwarding tree.

Fortunately, the above mentioned issues are currently the subject of on-going research [363] addressed by extensions to standard border gateway protocol (BGP) [364], termed as Multicast BGP. Multicast BGP extensions provides well established BGP-policy controls to specify RPF-check routing prefixes and thus, provide a finer grain of control for inter-domain multicast forwarding. What remains, before multicast is fully deployed in production IP networks as a ubiquitous routing service, is provisioning policies that can benefit *mutually* participating wireless ISP (WISP) domains.

To this end, it is interesting to observe that each wireless ISP is expected to be in control of some set of PoA clusters, serving a particular geographical location, in the form of a horizontal or vertical wireless network overlay.

Furthermore, HandoffCast is geared to serve as an L3-control protocol targeting IP mobility management of flows associated with *mobile* users. From this perspective, the demand for better tariffs or connection quality combined with user mobility, is expected to strengthen *cooperation* between multiple WISPs, towards cooperative support of delay seamlessness, for users in need of interactive IP services on the move.

Mobile users can, thus, be expected to traverse over multiple ISP domains maximising their benefit utility over multiple temporal carriers of their communications flows, within a certain geographical location. The former departs from the traditional

[29] or revenue

multicast service model of application services forwarding where usage is bound by the availability of interested but *static* users.

We argue that the above can act for WISPs, as an incentive of meshing their PoA infrastructures on the edge of their domains (through inter-domain routing), in a unified mobility management service. The latter would act, through HandoffCast, for all participating WISPs, both as a revenue accrual avenue, as well as a infrastructural provisioning unifier of ubiquitous, proactive but above all seamless mobility management service.

We may, thus, conclude that the usage of multicast for IP mobility management purposes, in a mesh of wireless network overlays, provides promising incentives of deployment amongst all WISP irrespective of their size. These incentives stem from guarantees towards a proportional share of priced network utility, for WISPs within the wireless Internet last-hop infrastructure, readily consumable by mobile users. The former encourages a proportional opportunity to multi-lateral pricing/provisioning policies amongst WISPs. Hence, the function of multicast coupled with the proportional share of WISP utility, emerging as a result of the user mobility, encourages shared pricing policies among all WISPs multiplexed onto a unified last-hop wireless Internet.

6.3 Conclusions

This chapter presented a comparative analysis between proactive MIPv6 and the emerging Fast handoffs proposal of the IETF.

From the analysis we have concluded that in its current instantiation, the FMIPv6 proposal is found to be hindered by a number of performance issues, over which PMIPv6 appears to perform significantly better.

The handoff delay performance of FMIPv6 is found to be sensitive to a significant NCoA/NAR miss probability during PoA determination for MN's next IPv6 handoff. The latter cascades into incorrect tunnel establishment and thus, recurrent black-hole effects and associate packet loss, in cases where multiple AP candidates are detected during an active PoA scan.

Furthermore, the signalling efficiency of PMIPv6 appears to be significantly better from both the perspective of the MN and AR, both for IP Roaming state establishment as well as forwarding during the critical period of the IP handoff.

PMIPv6 does not impose a dependency on link-specific event signalling or router discovery, when proactive IP-Roaming state establishment must be effected, as is the case with FMIPv6. Instead, PMIPv6 resolves all required state with a single control signal after MN's handoff has completed.

FMIPv6 does not specify accurately a viable tunnel activation method. On the contrary PMIPv6 guarantees instant activation of HandoffCast with no reliance on timeout periods.

Furthermore, FMIPv6 effects a DAD handling approach that would incur either the generation of additional state, additional control signalling or ultimately a delay measure similar to reactive DAD delay met in Mobile IPv6. On the contrary, PMIPv6 imposes no such performance overheads during the period of MN's handoff; the MN simply requests the establishment of IP-Roaming state at the previous PoA which in turn delegates the request to its PoA neighbours. Generation of MN's sCoA (NCoA) address at each candidate PoA is performed through proxy-stateless DAD resolution on-link, by means of standard Neighbour Solicitation. Since the proxy-stateless address configuration function is performed well in advance of MN's next IPv6 handoff, the standard DAD delay of 1000ms does not affect MN's packets communications during a PMIPv6 handoff.

Chapter 7

Contributions and Future Research Directions

7.1 Contributions

Reactive mechanisms in IPv6 mobility management standards, with respect to the MN's IPv6 handoff, are insufficient for the support of interactive IP application services. The deficiency pertains to reactivity in current functions of MN registration/association with its new point of attachment (PoA), supported by current IPv6 mobility management standards. Such deficiency introduces significant IP handoff delay and associated packet loss that becomes exacerbated as the MN experiences increased handoff rates. Mobile node registration or association with a single PoA has been defined as the process of state establishment of one (or more) contexts pertaining to MN's continuous IP link connectivity.

The central goal of this thesis has been to investigate factors that can impede delay-transparent performance in interactive applications, as a result Mobile IPv6 support. In view of these factors, this study has reconsidered the architectural framework set by the Mobile IPv6 standard and its derivatives. To this end, this thesis has looked into advanced, forms of IP handoff and flow forwarding management in support of delay-seamlessness, for interactive IP application services as well as capability-based norms for intelligent handoff control. The core contribution of this investigation may be summarised as follows:

By means of proactive registration to immediate candidate points of IP attachment, the handoff and flow forwarding management task can support realistically delay seamlessness in support of interactive IP application services. As a result, of proactive capability-based signalling IP mobility management can provide a sound basis towards intelligent handoff control

Our investigation encompassed a detailed set of experimental measurements, exposing performance short-comings in current reactive Mobile IPv6 standards, during an IP handoff. To this end, we devised a Mobile IPv6 experimental testbed supporting VoIP communications between the MN and its fixed CN/HA peers. As part of our experimental methodology we separated traffic monitoring at two distinct levels: (i) the network and (ii) the link layer where a handoff is manifested. Results derived looked at three individual signalling areas pertinent to handoff performance: (i) core IPv6 control signalling, (ii) core MIPv6 control signalling (iii) core link-layer (802.11) signalling. We identified and exposed intrinsic influences between these classes of signalling pertinent to the overall process of handoff handled by the current Mobile IPv6 standard. Findings from this investigation provided the basis of subsequent study reconsidering the architectural foundations of reactive norms in IPv6 mobility management

The following sections provide a summary of our contributions, discussing possible limitations as well as alternative approaches to addressing the problems at hand. To this end, we look at possible next steps of future research efforts.

7.2 Identifying performance strengths and trade-offs

In Chapter 2, we focused on an in-depth analysis of standard as well as most prevalent mobility management proposals. In this investigation, our contributions have been:

- We engaged into a signalling analysis of important mobility management mechanisms (see Annex C.1) and presented the benefits and design trade-offs involved. We found that simplistic protocol design becomes an unavoidable trade-off in handoff performance. It appears that the signalling of simplistic mobility management mechanisms is significantly influenced by external delay factors such as end-to-end delay. This introduces unwanted delay components in the handoff performance of the mobile node (MN).

- We concluded that, global location management signalling in IP network infrastructures, does not meet the limitations of cellular networks; this is because IP networks can massively over-provision the Internet backbone to deal with packet switched signalling that gets eventually distributed to autonomous systems. Autonomous systems support higher routing path redundancy that can accommodate the routing of such signalling. On the contrary, legacy cellular systems have very little routing redundancy in the face of signalling congestion. At the same time, over-provisioning the cellular network core with similar bandwidth to that of the

Internet backbone becomes prohibitively expensive. For this reason, we have concluded that micro-mobility mechanisms, although appealing in terms of observed inter-domain locating update savings, may not significantly contribute to IP mobility management performance, other than the reduction of signalling round trip times.

- We have identified that RTT latency savings, become a function of the network domain size and the rate of inter-domain mobility (see Section 2.3.3). For very large network domains an RTT of 50-150ms is not statistically uncommon; in such cases delay reduction must be balanced with system resiliency. For multi-overlay network infrastructures inter-domain vertical handoffs are bound to be the norm rather than the exception. In such cases, we showed that inter-domain handoff incur *more* signalling to micro-mobility protocols than their macro-mobility counterparts.

- We have shown that micro-mobility protocol mechanisms introduce significant complexity in terms of failure resiliency as well as localised mobility agent configuration and distribution over network partitions within an administrative domain (see Annex C.3). We find that micro-mobility protocols are expected to incur either:

  - extensive changes to the existing IP routing infrastructure
  - increased routing state on IP routers in the form of host routes
  - additional configuration mechanisms for load balance of traffic over multiple mobility agents (LMAs) distributed at the edges of the network domain
  - sub-optimal routing in cases where multiple border routers identify multiple edge routing paths to the Internet backbone. This is incurred by the fact that the MN registers with one LMA at one edge routing path, while traffic arrives to the MN from a different (edge) routing path.

Despite their performance trade-offs, simplistic macro-mobility management mechanisms such as Mobile IPv6, can be increasingly scalable with little effect on the network infrastructure. This is clearly not the case for micro-mobility protocols. The benefits introduced in the micro-mobility management function of dominant proposals are questionable, since operational scenarios reveal the need for additional mechanisms to sustain localised mobility management savings, originally praised under ideal operational conditions.

The above lead us to conclude that macro-mobility management design strategies are to be preferred over micro-mobility techniques. Clearly, this is in line with the doctrine of Fast handoff extensions for Mobile IPv6. FMIPv6 attempts to augment the mobility management function to accommodate provisions for delay seamlessness acceptable for interactive real-time applications. While the proposed extensions are still under investigation, emerging results from independent investigations on the performance of FMIPv6, report prohibitively large handoff delays as a result of an increased number of wireless hosts attached to a wireless link. Reason for this pertains clearly to the dependency of FMIPv6 signalling on access contention before signals can be propagated between previous and new ARs.

From the perspective of *flow forwarding* we find that:

- two viable alternatives may exist: (i) forwarding over a unicast tunnel (ii) forwarding to a multicast group. Solutions of multiple tunnels impose replication of forwarded traffic onto multiple paths. Multiple tunnels to remain scalable have to remain fixed between old and new ARs. This presents the additional complexity of identifying individual flows for the particular mobile node; once the flow is de-tunnelled at the new AR, the packet has a topologically incorrect destination (the previous CoA of the MN valid at the old AR). As a result, nested (pair-wise) encapsulation is required for the AR to distinguish the destination MN at the new point of attachment once the outer header has been removed.

- for unicast tunnelling at the new CoA, the significant limitation appears to be the accuracy of resolution of the MN's prospective new CoA. If the latter is not accurate, MN's traffic will be forwarded to the wrong PoA, reinstating the transient black holes - in terms of packet loss - that Mobile IPv6 introduced. On the contrary, forwarding to a multicast group appears to be free of such limitations. However, it becomes apparent that multicast forwarding cannot be effected at the peers of the MN, that is, HA or CN as is the case with the Mysore et al and DAEDALUS proposals; this is because the multicast tree management becomes prohibitively expensive. On the contrary such cost becomes more manageable when the multicast tree is emanated within the network domain visited by the MN. This is the case with the IDMP and M&M proposals.

- the fundamental limitation of multicast forwarding, when rooted at the *edge* of a network domain, is the classic limitation of micro-mobility protocols: a single point of failure can collapse the operation of the mobility management protocol across the entire domain. In cases where the RP is co-located at the edge LMA, a node failure implies also total failure of the flow forwarding mechanism. Where the RP is not co-located with edge LMA then in the case of an LMA failure the flow forwarding mechanism requires an additional LMA discovery and re-election mechanism to direct all traffic to the RP, assuming that all traffic on the failed edge can be successfully routed to the back LMA. In any case, in such situations flow forwarding is guaranteed to experienced sub-optimal routing, as well as increased end-to-end delay. The latter may impact the delay performance of interactive communications between the MN and its peers.

Both current unicast and multicast forwarding approaches suggest that a more robust and distributed flow forwarding mechanism is required. It is imperative that such mechanism can sustain failure of the forwarding node without destructive effects in flow redirection to all MNs in the network domain.

7.3 Performance shortcomings of MIPv6

In Chapter 3, we presented a detailed set of experimental measures regarding VoIP performance over Mobile IPv6 enabled Wireless LAN communications. From the measurement traces collected and subsequent statistical analysis we identified a comprehensive set of findings contributing to a detailed model of Mobile IPv6 behaviour, in terms of delay performance. In particular:

- Duplicate address detection incurs a significant amount of delay consistently for all handoffs measured.

- We have shown that a high rate of router advertisements[1] *does not*, by itself, assure a faster handoff (see Annex D.5.1). We have shown that in best cases a minimum average of 80ms of hangover delay arises from the moment the L2-handoff has completed until any subsequent neighbour discover function. The fundamental reason is that movement detection on the MN relies two possible time-based events: (i) the expiry of the router advertisement lifetime (ii) untimely initiation of router solicitation.

[1] average router advertisement interval reduced to 50ms

- We have shown the timing of a router solicitation during a MIPv6 handoff is found to be inappropriate, in a way that it impacts significantly the movement detection process (inducing the sap-reactive hangover delay component) and thus, impedes the prompt completion of the MIPv6 handoff process.

- the rate of router solicitations during movement detection was found to be such that impedes an expedient movement detection. As a result the MIPv6 movement detection delay impacts in turn, the handoff process, incurring significant delay.

- Response to router solicitations, excluding the one that originated during an L2-handoff in progress was found to consume around 50-60ms, before a unicast router advertisement is sent to the MN.

- An authenticated binding update to a single corresponding node is found to incur significant delay that is guaranteed to increase the delay of MIPv6 handoff (see Section 3.7.1). For statistically derived worst case RTT scenarios this ranged up to 484ms in the path between the MN-HA, MN-CN or HA-CN host pairs.

- The delay of a MIPv6 handoff for the MN return back to the home network (v2h MIPv6 handoff) is found to be significantly smaller (see Section 3.7.1) than the respective delay incurred during a handoff to a visited network (h2v MIPv6 handoff). Nonetheless it remains excess of 200ms and thus capable of impeding interactive delay performance.

- The significantly smaller measure of MIPv6 handoff delay on an v2h MIPv6 handoff is accounted by the existence of the HA entity co-located on the same host as the Access Routing function. This implies, that the MIPv6 protocol effectively requires the existence of the HA function at the router devices to allow for low MIPv6 handoff delays incurred by the Mobility-enabled IPv6 layer[2].

- The delay imposed by the neighbour unreachability detection process is significant (see Section 3.7.1), so as to impact receipt of the Binding acknowledgement from the HA as well as the subsequent update of bindings (with or without authentication) at the communicating peers.

- Hangover delay cause during movement detection (see Section 3.20) is found to contribute a significant delay component to the total MIPv6 handoff latency.

[2]that is excluding the L2 handoff delay which is owed to delays incurred by the wireless technology.

While this may be possibly accounted by inefficiencies in the implementation of the IPv6 specification (i.e. neighbour discovery), it attests that for different mobile devices, potential inefficiencies (induced heterogeneity), dependence on IPv6 neighbour discovery signalling has a cascading effect on the total MIPv6 handoff delay component.

- From the perspective of packet loss for a packetisation rate of 20ms and no suppression of silence packets, an h2v handoff was found to incur loss runs above between 150-154 packets. While voice activity detection (VAD) can reduce the amount of packet loss within the same period of handoff delay, the amount of reduction is dependent on the degree of activity of the voice conversation between the participants or the composition of a voice conversation. For experimental purposes VAD-enabled VoIP flows are left as future work for the purposes of this investigation.

- MIPv6 is found to incur significant jitter during a MIPv6 handoff (see Section 3.8). The measure of jitter induced is accounted by actual packet loss during the MIPv6 handoff, and is capable to offset significantly lip synchronisation in potential interactive multimedia that comprise of both audio and video. The measure of the MIPv6 handoff and imposes a minimum jitter amortisation period before which any subsequent handoff will simply worsen the amount of jitter experienced as well as the loss of lip synchronisation .We remind that these packets cannot be recovered by either increases play-out delay at the receiver or by packet loss concealment techniques

With respect to delay incurred at the link layer of IEEE802.11, as the wireless technology of choice during this set of experimental measurements, we found that link-layer delay is *independent* of the MIPv6 handoff delay observed at the IPv6 layer. At its current form we concluded that it affects the prompt completion of a MIPv6 handoff in the following ways:

- The MAC layer of the 802.11b WLAN specification is found to incur significant delay (see Section 3.7.1) which by itself can impede any guarantees of real-time delivery for VoIP packet flows between the MN and its peers.

- The dominant delay component in an L2-handoff is incurred by the AP-discovery phase, as the MN scan reactively the channels triggered by a low SNR threshold.

- During an L2-handoff *all* channels are scanned before an AP is selected (see Annex D.6). While the algorithm of AP-discovery and in particular the ordering in which channels may be scanned is not mandated by the protocol, it is clear that the scanning process remains agnostic of the candidate APs surrounding the MN, unless explicitly scanned *in reaction* to signal loss below a certain SNR threshold.

- For an increasing number of associated MNs under the same AP (see Annex D.7), any reliance of the MIPv6 to small measures of router advertisement interval was found to be detrimental to the completion of the movement detection process and as a result, the completion of the MIPv6 handoff. This is because the size of the advertisement interval is offset by a significant amount of frame delay which effectively results into a prolonged movement detection before the first router advertisement is received.

It may be seen that even by ignoring the L2-handoff delay component incurred by the MAC layer of 802.11b, the delay incurred by the MIPv6 layer alone is sufficient to place a VoIP flow below any guarantees of interactive real-time delivery of IPv6 traffic.

The above provide overwhelming evidence that the MIPv6 protocol standard incurs significant delay at the IPv6 layer, such that it cannot preserve the seamlessness principle so as to sustain real-time guarantees in the delivery of interactive multimedia services to wireless IPv6 mobile devices.

While in both v2h and h2v handoff cases there exists a significant L2 handoff delay components of about 420ms which is beyond the control of MIPv6 mobility management, the MIPv6 mechanism:

- experiences significant delays at the IPv6 layer of the network stack on handoffs away from the home network.

- remains agnostic of the link-layer mechanism at the cost of a significant L2 handoff delay component and makes no provisions that can alleviate such delay component, characterising the total MIPv6 handoff delay measure.

Furthermore, as part of our contributions we have:

- we have also derived a number of statistical distributions for the individual delay components experienced during a MIPv6 handoff (see Annex D.4 and D.5). This set of statistical distributions are used in subsequent simulations to describe stochastically the statistical measure of MIPv6 handoff delay during simulations, when compared with any novel IPv6 mobility management proposed in later sections.

- we have shown that true MIPv6 handoff performance is conflicting with performance claims made by Mobile IPv6 specification 3.7.1. Mobile IPv6 *cannot* guarantee a handoff rate of 1 handoff/sec; a MIPv6 handoff away to a visited network (h2v) is guaranteed to last a minimum of 2.6 sec. Our results show that the Mobile IPv6 specification should align its claims of handoff rate support to a rate of a handoff every 3 seconds 0.35 h/sec), without including any influence for network externalities such as RTT disparity over Internet links.

7.4 Proactive IPv6 Handoff Management

In Chapter 4, we reconsidered the architectural foundations of Mobile IPv6, by investigating novel Mobility management mechanisms that support by design delay transparency. To this end we proposed Proactive IPv6 mobility management. Proactive handoff management is found to eliminate reactive handoff delay arising at the network layer, by:

- identifying the immediate handoff AR neighbours (see Section 4.7) with respect to MN's current point of attachment through Handoff AR discovery (HARD).
- promoting state establishment pertinent to MN's (IPv6) network connectivity at the next PoA, well in advance of its IPv6 handoff transition.
- establishing a sufficiently abstract mapping between the network and the link layer that allows expedient movement detection at the network layer with no reliance on network layer signalling (such as router advertisements).
- ensuring that cross-layer optimisations are abstract and generic enough to be feasible/available across all wireless technologies.

Core contributions arising from this study may be summarised as follows:

- we identified a robust mechanism for proactive state establishment (see Section 4.8). We found that for these purposes the MN must remain associated with it current PoA for a minimum cell residence period. The measure of this period is dependent on the complexity and the delay requirements for state establishment/evaluation at the HAR neighbour.

- we proposed and evaluated a mechanism for discovery of MN handoff AR candidates (HARD). There we showed that the abstraction between a Routing (network-layer) and a Mobility (link-layer) Neighbourhood enables the MN to transform physical node movement into mobility-hop Roaming state. Such transform can effect a proactive IPv6 handoff with no dependence on traditional functions of IPv6 neighbour discovery such as Address Resolution or router advertisement signalling.

- we have identified a generic set of cross-layer optimisations (see Section 4.8.3) to support robust initiation or termination of core proactive handoff management functions. Such optimisations utilised a generic set of *AP identification information* available at the link layer, in the form of triggers for accurate handoff management. Such type of information can be safely generalised for *any* wireless networking technology, since APs must remain at all times identifiable for management purposes.

- by combining the above we have identified forms of reducing (i) dependence on core IPv6 signalling (ii) the measure of core IPv6 signalling required and (iii) its associated probability of MAC contention.

- we identified and evaluated (see Section 4.11.1) a mechanism that increases MN's service utility by allowing PoA diversity based on selection policies during the IP handoff decision.

From the performance analysis of proactive versus standard reactive handoff management we concluded that proactivity handoff management *can* address successfully delay seamlessness at the *network* (IPv6) layer.

However, looking at the total handoff delay measure, we found that proactive handoff delay is well above the 200ms requirement imposed by interactive real-time services. Thus, while proactive handoff management *can* eliminate delay incurred by network layer functions during an IPv6 handoff, it is found to be insufficient *by itself* to address the *total* handoff delay.

The measure of delay incurred in proactive handoffs is owed primarily to two latency factors over which the network layer can exert no control: (i) the link-layer handoff delay, (ii) round trip time delay. In the observed proactive handoff delay measure, 380-420ms of the delay is owed to the link-layer handoff, while 80-100ms to the average round trip time delay.

We have indicated further that, different WLAN implementations and in general different wireless technologies achieve different L2-handoff latencies. For instance, certain WLAN implementations can achieve L2-handoff delays as low as 150-160ms.

Assuming a hard delay bound of 200ms for interactive real-time services and 80-100ms as the dominant measure of round trip time observed during the simulation study, it appears that provision of any realistic guarantees towards delay seamlessness in proactive IPv6 mobility management requires a maximum L2-handoff delay bound of $\leq 100ms$. In this manner, the combined upper bound of these two delay components may thus, be attacked through proactive MIPv6 management by ensuring that MN's traffic is also *redirected* towards it new PoA while a proactive handoff is in progress.

7.5 Seamless flow forwarding through HandoffCast

In Chapter 5 we investigated the performance of HandoffCast as a proactive managed flow forwarding mechanism. HandoffCast complements the overall Proactive IPv6 mobility management task towards support of delay seamlessness.

From a detailed set of simulations conducted we showed that HandoffCast forwarding *is* capable of addressing successfully delay seamlessness during MN's IP handoff. Such capability is, however, directly related to specific performance factors, identified and optimised throughout this simulation study. Our contributions in this study may be summarised as follows:

- We showed that these such factors stem from two fundamental delay sources that remain beyond the explicit control of the network layer: (i) the measure of link-layer (L2) handoff effected by the individual wireless technology at hand, (ii) the measure of one-way delay arising from the perspective of flow forwarding during MN's IP handoff, in an effort to eliminate black-hole effect and reduce the measure of perceived flow disruption.

- We identified that HandoffCast performance is critically dependent on the measure of the delay contributed by the standard 802.11 L2-handoff process. To this end, we have proposed an L2-handoff delay optimisation that reduces drastically the latency incurred during the critical period of active PoA scanning with no negative impact to function robustness.

- We showed by means of simulations that HandoffCast forwarding delay performance, appears to be below the 200ms threshold, but remains dependent on the placement of the multicast RP (see Section 5.6.1). Our simulation results suggested that RP placement at an AR with a *high* node-degree ($\geq$ 6) sustains *scalable* HandoffCast forwarding delay performance for increasing PoA densities. On the contrary, for an RP node-degree of $\leq$ 3, the measure of persistent delay grows significantly on the forwarding path (old PoA $\rightarrow$ RP $\rightarrow$ new PoA). Such effect is exacerbated when considering the total e2e delay over leaf-PoA densities $\geq$ 140, between the CN $\rightarrow$ old PoA $\rightarrow$ RP $\rightarrow$ new PoA

With respect to L2-handoff delay performance over 802.11 WLANs, during HandoffCast forwarding, we attained an average measure of L2-handoff delay that does not exceed 60-70ms for small or large PoA densities (i.e. 45-200 leaf-PoA nodes) with the aid of proposed optimisations (see Annex F.5). Such L2-handoff delay measure approaches the level of L2-handoff over cellular networks.

Given a random allocation of channels among members of an R-neighbourhood that comprises of different WISPs, we showed that the average number of operational channels within an R-neighbourhood is normally distributed with a mean of 6 (see Section 5.6.1). For an increasing R-neighbourhood size, the variance of the number of operational channels appears to maintain an upper bound of 70% of the R-neighbourhood size (c_n), for $c_n \geq 10$.

The above results (see Annex F.5), confirm that proactive guiding of the AP scanning process over channels that are operational with MN's R-neighbourhood can: (i) reduce L2-handoff delay significantly in a realistic operational scenario, (ii) support a low measure of persistent delay, capable of addressing delay seamlessness during MN's IPv6 handoff, (iii) avoid energy-intensive and packet-loss prone *interleaving* [335, 336] between transmission and AP scanning modes in search of available PoA.

7.6 Performance of Proactive versus Fast MIPv6

in Chapter 6 we presented a comparative analysis between proactive MIPv6 and the emerging Fast handoffs proposal of the IETF. Contributions from this investigation encompassed:

- a parametric model of three-cell overlap (see Section 6.2.2) essential for signalling analysis through simulation.

- the identification of the NCoA/NAR miss probability metric for the purposes of assessing the accuracy of NAR determination in fast handoff proposals (see Section

6.2.1).

- a signalling analysis between the two protocol proposals (see Section 6.2.3).

From the analysis we have concluded that in its current instantiation, the FMIPv6 proposal is found to be hindered by a number of performance issues, over which PMIPv6 appears to perform significantly better.

In particular, the handoff delay performance of FMIPv6 is found to be sensitive to a significant NCoA/NAR miss probability during PoA determination for MN's next IPv6 handoff. The latter cascades into incorrect tunnel establishment and thus, recurrent black-hole effects and associate packet loss, in cases where multiple AP candidates are detected during an active PoA scan.

Furthermore, we found that the signalling efficiency of PMIPv6 appears to be significantly better from both the perspective of the MN and AR, both for IP Roaming state establishment as well as forwarding during the critical period of the IP handoff.

We found that PMIPv6 does not impose a dependency on link-specific event signalling or router discovery, when proactive IP-Roaming state establishment must be effected, as is the case with FMIPv6. Instead, PMIPv6 resolves all required state with a single control signal after MN's handoff has completed.

We have shown that FMIPv6 does not support a fast tunnel activation method (see Section 6.2.5 and 6.2.6); as a result it can reintroduce reactive mobility latencies and associated packet loss. On the contrary PMIPv6 guarantees instant activation of HandoffCast with no reliance on timeout periods.

7.7 Critical Review and Future Research Directions

This section provides a brief elaboration on the assumptions made in parts of this investigation and contemplates certain design issues and technical decisions taken. To this end it proposes a number of suggestions to mitigate potential limitations and present options for future work.

7.7.1 Simplifying assumptions and potential limitations

During our investigation by means of simulation, a number of simplifying modelling assumptions were made. In particular, the propagation environment was assumed to be free-space and unobstructed; the AP coverage area was assumed to be circular; random way-point movement was assumed as the mobility pattern of the MN; small transmission ranges were modelled in proportion to the movement grid.

We revisit these assumptions and their potential limitations showing new avenues of future research directions in the evaluation of the proactive MIPv6 architecture.

Wave propagation in real-world wireless environments is not free-space nor propagation environments are unobstructed. Terrain obstructions or foliage can induce multi-path fading or loss on the energy of the propagated wave. Such effects can increase the bit error rate (BER) over the air interface by corrupting communicated frames at the link layer. However, corrupted transmissions as a result of propagation effects are typically managed by means of power control (temporal increase of power at the transmitter) and/or by retransmission and error correction techniques, effected over the air interface on sub-millisecond basis at the link and physical sub-layers. Depending on the harshness of the propagation environment the BER metric may exhibit small (i.e. negligible) fluctuations that does not affect the packet communication flow or large (i.e. significant) fluctuations that impacts packet transmission at the network layer. This work has looked at the performance of interactive real-time services over the a novel mobility management architecture, whereby the wireless interface exhibits negligible BER fluctuations by assuming open space unobstructed propagation environments. From the perspective of IP mobility management performance over handoff and flow forwarding management signalling, an interesting future research direction would be to investigate the measure of influence of high BER fluctuations over harsh propagation environments. It would be interesting to quantify the robustness of signalling acknowledgement and under what conditions (e.g. high offered load emergencies) additional optimisations may be required.

With respect to the assumption of circular AP coverage areas, we identified in Section 4.9.3 that for cross-polarised transmit/receive environments the coverage area is not necessarily circular but lobe-shaped. We have argued however, that cell shape is not expected to affect the performance of mobility management mechanisms from a BER perspective; this is because the mobility mechanism at hand depends on the availability of some *minimum measure of overlap between coverage areas (whether circular or lobe-shaped) rather than their shape*; otherwise dis-connectivity at the link-layer should be expected. The only influence that cell shape may have on the performance of mobility management mechanisms is on the measure of handoff rate. Irregular cell shapes may affect the measure of handoff rate as well as the number of ping-pong effects depending on their placement with respect to MN movement path. For proactive MIPv6, variation of the handoff rate may affect the rate of HARD state convergence and the number of handoffs per interactive communication session. To this end, future research direction would encompass the influence of irregular AP cell shape on the measure of handoffs during an interactive communication session.

With respect to the mobility model adopted, this work looked at the behaviour of random way-point MN mobility patterns. However, in actual terrain mobility environments MN movement is not necessarily random; it is bound by particular obstructions such as buildings or hills as well as the placement of APs with respect to these obstructions. Furthermore, the number of changes in direction, available to the MN at each way point is not always small (see Section 4.9.4); for instance under the circular or Manhattan mobility models, the number of available changes in directions may be larger (1-6). Future work will be looking at the influence of alternative MN terrain mobility models, encompassing building obstructions, on the performance of the proposed mobility management architecture. Particularly interesting is expected to be the aspect of adaptation of the respective wave propagation model, to reflect the obstructions emerging under either Manhattan or circular movement terrain models. Use of different mobility models is expected to influence the handoff rate per communication session by either increasing[3] or decreasing[4] MN's cell residence period.

It should be also noted that for most part of this investigation simulations or experimental measurements have adopted the IEEE802.11 WLAN specification as the wireless interface. Main reason for such choice is the pronounced measure of L2-handoff delay over such wireless technology in contrast to cellular networks.

7.7.2 Experimental results and limitations

The experimental results of MIPv6 delay performance attained in Chapter 3 are based on a well-known, nonetheless *particular*, implementation of Mobile IPv6. During our investigation we have acknowledged that the sap/numb-reactive hangover delay component (see Section 3.7), cannot be clearly attributed on the actual performance of the movement detection of MIPv6 or an implementation inefficiency. Such ambiguity may be attributed to:

- lack of a protocol standard function for MIPv6 movement detection purposes, which also accounts as a shortcoming in MIPv6 delay performance.
- implementation inefficiencies that causes the MN to remain insensitive to Router Advertisement during the period of hangover delay; this may be the result of an

[3] introduction of obstacle-based changes in direction and as a result decrease in MN speed
[4] removal of obstacles with constrains in direction and increase in MN speed

inconsistent or erroneous router advertisement timer implementation that is not strictly adherent to fast resolution.

To this end, it is important to acknowledge that the reported MIPv6 delay performance may have been influenced to a *small* degree by implementation inefficiencies or issues, while exposing potential MIPv6 protocol/implementation shortcomings. Nonetheless, the degree of influence from potential implementation inefficiencies remains small; the measure of hangover delay is significantly smaller than the measure of delay incurred by DAD and NUD that are explicitly protocol specific. Future work would revisit the issue of hangover delay over a commercial MIPv6 implementation and contrast (as well as validate) its delay performance components against the HUT experimental implementation.

7.7.3 Qualitative performance of VAD-assisted VoIP flows

With respect to interactive service performance over reactive or proactive MIPv6, the results presented in Chapters 4 and 5 have been based on constant bit rate (CBR) VoIP flows with no voice activity detection. Due to the CBR character of the VoIP flows explored and given that the correspondent node remained fixed during communication, our measurements focused on unidirectional component flows of the VoIP conversation. This is because the pattern of both latency and associated loss was found to be identical at the CN. Under such setting the measure the measure of handoff delay, loss and jitter identify provide an upper bound of IPv6 mobility management performance.

However, in a commercial setting where capacity over the wireless last hop is a limited resource it is expected that VoIP sessions would utilise voice activity detection functionality, for the purposes of bandwidth savings. Under such scenarios a VoIP flow would exhibit a variable bit rate pattern, where silence packets are suppressed. To this end, we are interested in looking at MIPv6 handoff delay performance during VAD-assisted (VBR) VoIP flows. Under such experimental scenarios both unidirectional VoIP flows need to be monitored, since their pattern is asynchronous and asymmetric.

It is expected that under such communication scenarios MIPv6 handoff delay performance, interactive IP application flows would experience unidirectionally a smaller measure of both handoff delay and packet loss. To get a coarse measure of such delay we recall from Section 3.3 that voice may be approximated by considering the characteristic measures of talk-spurt (on) and silence (off) periods from actual VoIP session. Such periods may modelled as exponentially distributed interval with mean 352ms (talk-spurt) and 648ms (silence) respectively.

Statistically these mean values may also represent a probability of 0.35 of voice content and 0.65 of silence during the period of MIPv6 handoff delay. By looking the performance of the talk-spurt probability within, or at the boundaries of the measure of MIPv6 handoff delay we can derive an expected measure of MIPv6 handoff delay affecting VAD-assisted VoIP sessions. In particular, assuming a talk-spurt probability of 35% over a MIPv6 handoff delay of 2.8 sec, it is easy to see that the expected delay incurred by a full talkspurt occuring within this period is of the order of around 980ms. Such delay measure translates to nearly 49 packets lost (assume 20ms packetisation rate) during a handoff. To account for the talkspurt occuring at the boundaries of the handoff delay, we may approximate the measure of delay incurred by considering half of the original talkspurt probability (i.e. 17%). In such case, the delay incurred would be 476ms or nearly 24 packets lost during a handoff.

To strengthen the validity of this brief analysis, we consider further an ideal MIPv6 implementation, without hangover or neighbour unreachability delay effects and zero-time acquisition of router advertisements. Under such ideal operational scenario the sole source of delay is assumed to be DAD delay, L2-handoff delay and round trip delay. Assuming that for these 3 factors the measure of MIPv6 handoff delay is 1.5 sec and by applying the talk-spurt probability within this delay period, we find that a MIPv6 handoff experiences a minimum of 525ms and a loss run of 26 packets. In the event that the talk-spurt period falls at the boundaries of the handoff period, the minimum expected measure of MIPv6 handoff delay appears to around 255ms with a loss run of 13 packets.

It is important to note that in *all* above scenarios the expected measure of MIPv6 handoff delay is the *minimum* since we do not account for the number of on-off occurrences within a MIPv6 handoff delay period.

Without entering a more elaborate analysis by either experiments of simulations, it is easy to see that VAD-assisted VoIP communications experience a significant measure of delay during a MIPv6 handoff. Despite the above, we remain interested in exploring the deviation of experimental results from the above coarse expected measures of MIPv6 handoff delay.

Alternative direction to this research is to expand our performance analysis towards objective or subjective quality measures of VoIP services of IPv6 mobility management mechanisms.

7.7.4 Optimising the size of the PoA R/M-Neighbourhood

To optimise the signalling cost incurred under PMIPv6's pessimistic mode of operation in the management of PoA receivers during flow forwarding, PMIPv6 prescribes further an *optimistic* mode of handoff AR determination; in this, the full set of AR neighbours is reduced to a minimal subset of AR neighbours that exhibit the highest probability ranking for handoff candidacy. Such mode of candidate AR determination does not aim to detect precisely the identity of a single AR handoff candidate, as pursued by FMIPv6. To this end, future work will be looking at investigating the effect of predictive techniques on the measure of PoA neighbours, exhibiting prime candidacy for MN's next handoff.

Appendix A

Glossary

This Appendix provides a glossary of all acronyms and abbreviated terms used during the course of this thesis.

Abbrev.	Technical Term
AR	Access Router
BA	Binding Acknowledgement
BSS	Basic Service Set
BU	Binding Update
CAP	Current Access Point
CoA	Care-of Address
CoT	Care-of Test
CoTI	Care-of Test Initiate
CN	Correspondent Node
DAD	Duplicate Address Detection
DHCPv6	Dynamic Host Configuration Protocol for IPv6
ESS	Extended Service Set
FBA	Fast Binding Acknowledgement
FBU	Fast Binding Update
FNA	Fast Neighbour Advertisement
HA	Home Agent
HACK	Handover Acknowledgement
HoA	Home Address
HoT	Home Test
HoTI	Home Test Initiate
HI	Handover Initiate
MN	Mobile Node
NA	Neighbour Advertisement
nAP	New Access Point
nAR	New Access Router
nCoA	New Care of Address
NeighSol	Neighbour Solicitation

Abbrev.	Technical Term
pAP	Previous Access Point
pAR	Previous Access Router
pCoA	Previous Care of Address
PrRtAdv	Proxy Router Advertisement
QoS	Quality of Service
RA/RtAdv	Router Advertisement
RR	Return Routability
RS/RtSol	Router Solicitation
RtSolPr	Router Solicitation for Proxy
WLAN	Wireless Local Area Network

Appendix B

Wireless Technology Task Groups

Section B.1 provides a short description of new and emerging technologies populating commercially the air interface.

Section B.2 presents a brief description of important task groups, current active in the arena of wireless technologies. In addition, it provides a brief view of non-IEEE wireless technologies receiving increasing popularity in the domain of IP networks.

B.1 New and emerging Wireless Technologies

Currently, the space of wireless communications is segmented in four distinct classes of wireless networks, each offering diverse performance characteristics. These are in the macro-cellular (satellite), the cellular (PCS/WiMax), micro-cellular (WLAN) and pico-cellular (Bluetooth/Zigbee).

With respect to macro-cellular environments, depending on the generation of satellite technology, wireless communications support aggregate bandwidths of 2-166 Mbps over a regulated frequency band. On the other hand, propagation delays may range from 20-100ms (LEO), 110-150ms (MEO) and 250-280ms (GEO), with national or multinational coverage footprints [365].

For cellular networks, the underlying cellular infrastructure uses a network of base stations (BS), each identifying a *cell* coverage area, from 1 to 50 km in radius over a regulated frequency range. Typical aggregate data rates over cellular networks can range, depending on the technology, from 120KBps (GPRS), 384Kbps (EDGE) up to 2 MBPS (UMTS)[1]..

During the last decade the cellular paradigm has been adopted for the creation of micro-cellular networking environments, aiming to support user mobility at smaller transmission ranges over de-regulated carrier frequencies. Numerous consortia and

[1]with corresponding (per-user) available data rates around 38 (GPRS), 70 (EDGE) and 100-384 Kbps (UMTS)

standardisation bodies have been working on independent and often non-interoperable micro-cellular network specifications. Today, more than 20 candidate technologies, working groups and standard specifications already exist; annex F.8 presents briefly important wireless technology efforts.

B.1.1 Micro-cellular environments: Wireless LANs

Amongst them, IEEE802.11 [17] is perhaps the most mature wireless protocol for wireless LAN (WLAN) communications, tested and deployed for years in corporate, enterprise, private and public environments, also posed as the favoured technology for many home networking applications.

The IEEE 802.11 standard supports several derivative 802.11 technologies in the unlicensed bands of 2.4 and 5GHz, while sharing a common medium access control (MAC) sub-layer over two physical (PHY) layer specification technologies: (i) direct-sequence spread spectrum (DSSS) (ii) frequency-hopping spread spectrum (FHSS).

Initially IEEE 802.11 systems operating at the 2.4GHz band, provided data rates up to 2 Mbps and propagation delays below 5 ms. However, its wide acceptance, initiated new revisions to the specification base giving rise to *high rate* (HR) extensions. The most important of these is the IEEE 802.11b PHY layer specification; 802.11b achieves aggregate signalling rates of 5.5 and 11 Mbps, using complementary code keying (CCK) modulation [51].

Recently, the IEEE 802.11g task group has formed a draft standard that achieves data rates higher than 22 Mb/s, adopting either single-carrier trellis-coded 8-phase shift keying (PSK) modulation or Orthogonal Frequency Division Multiplexing (OFDM) modulation schemes.

In the unlicensed 5GHz band, 802.11a [366] and HIPERLAN/2 [18] technologies support data rates up to 54 Mb/s at transmission ranges smaller than that of 802.11b; 802.11a achieves this using OFDM modulation, while HIPERLAN/2 through Time Division Multiplexing (TDM).

It is interesting to note that even standardisation by itself, is not guaranteed to produce interoperable implementations of a specified wireless technology. To this end, IEEE 802.11 product manufacturers have established an interoperability alliance identified as Wireless Ethernet Compatibility Alliance (WECA). Under this consortium, all participating partners ensure interoperability and compatibility of their 802.11 implementations through a suit of interoperability tests collectively identified as Wireless Fidelity (Wi-Fi) standard. Equivalent standards have been adopted for 802.11a implementations collectively identified as Wi-Fi5.

Along with the adoption of new modulation schemes in support of high data rate extensions to the 802.11 base specification, other 802.11 task groups have been formulating individual MAC or PHY protocol recommendations, aiming to enhance or supplement specific areas of the base 802.11 specification. A brief overview of these groups together with other competing technologies is presented in the following Section.

B.2 Wireless Technology Task Groups

A number of wireless technology task groups has emerged during the last decade. In an effort to provide perspective in breadth and depth of issues concerning wireless link performance, this section presents briefly, a set of important wireless technology efforts supported by individual standardisation groups.

802.11d 802.11d task group (TG) works towards 802.11b versions at other frequencies, for countries where the 2.4GHz band is not available.

802.11e 802.11e TG works towards the specification of a new 802.11 MAC protocol extensions, to support Quality of service provisions at the 802.11 MAC sub-layer. These extensions facilitate prioritisation of data, voice, and video transmissions by limiting the amount of back-off delay enforced for special types of multi-protocol data units (MPDU).

802.11f 802.11f TG aims to improve the handover mechanism in 802.11 so that users can maintain a connection while roaming between access points attached to different networks.

802.11h 802.11h TG aims to enhance the power control and radio channel selectivity of 802.11a, to ensure acceptance by European regulators.

802.11i 802.11i TG aims to address security issues arising from the current forms of security adopted by the 802.11 standard. Instead of the *Wired Equivalent Privacy* (WEP), a new authentication/encryption protocol based on the Advanced Encryption Standard (AES) is currently under investigation.

802.11j 802.11j TG aims to resolve HiperLAN/2 and 802.11a interoperability issues.

802.11k 802.11k aims to standardise mechanisms that allow a wireless local area network (WLAN) to perform channel selection, host roaming, and transmit power control (TPC) to optimise network performance.

802.11k is primarily intended to improve the way hosts and traffic is *distributed* within a network. Normally, a WLAN host device connects to the access point (AP) that provides the strongest signal. Depending on the number and geographic locations of the subscribers, this arrangement can sometimes lead to excessive demand on one AP and under-utilisation of others, resulting in degradation of overall network performance. In a 802.11k-conformant network, if the AP having the strongest signal is loaded to its full capacity, a wireless device is connected to one of the underutilised APs. Even though the signal may be weaker, the overall throughput is greater because more efficient use is made of the network resources.

There is a general disagreement in the networking community about the formal adoption schedule of this specification and subsequently how soon after any device implementation supporting it will become available. While working groups estimate deployment as early as 2005, vendors suggest that such process is still at its very early stages.

802.11u Referred to as the Wireless Inter-networking with External Networks (WIEN) Study Group, 802.11u is establishing standards for the integration of 802.11 and macro-cellular communication systems such as EDGE/UMTS. The task group is studying access router identification, MAC address anonymity, scalability, policy enforcement, access control, quality of service, and billing administration; in addition to the other requirements for inter-operation between network systems.

802.16a/e 802.16 is a group of broadband wireless communications standards for metropolitan area networks (MANs). The original[2] 802.16 standard, specified *fixed* point-to-multipoint broadband wireless systems operating in the 10-66 GHz *licensed* spectrum. Officially identified as Wireless MAN, the 802.16 specification is expected to enable multimedia applications with wireless connectivity with a range of up to 30 miles, providing an extended last-mile technology.

802.16a is a recent[3] a wireless communications specification for metropolitan area networks (MANs). Known as WiMAX specification, the 802.16a/e standards complement the older 802.11 family of specifications.

The 802.16a standard has been developed for wireless MANs operating on licensed and *unlicensed* radio-frequency (RF) bands between 2 GHz and 11 GHz, at data speeds of up to 75 Mbps, with low latency and efficient use of spectrum space. Security is enhanced by encryption features. Forward error correction (FEC) and space/time coding optimise accuracy under marginal signal conditions. The maximum range can be extended to approximately 30 miles (48 kilometres) with associated trade-offs in

[2]published in December 2001.

[3]approved in January 2003 and released in April 2000.

throughput. The 802.16a specification is ideally suited for advanced communication methods such as voice over IP (VoIP) and prioritised data traffic. 802.16e is expected to combine 802.16a and 802.20 capabilities with backwards compatibility for 802.11b/g under a single interface for mobile hosts.

802.20 The IEEE 802.20 WG aims to develop the specification for an efficient packet based air interface, optimised for the transport of IP based services.

Such interface fosters by design interoperable mobile broadband wireless access systems, operating in licensed bands below 3.5 GHz, with peak data rates *per user* in excess of 1 Mbps. It aims to support various vehicular mobility classes up to 250 Km/h in a MAN environment for cell ranges up to 15 kilometres.

802.20 targets sustained user data rates and numbers of active users that are significantly higher than achieved by existing mobile systems [241]. The 802.20 interface seeks to boost real-time data transmission rates in wireless MANs to transmission speeds similar to that of rival DSL and cable connections.

5GHz Unified Protocol

IEEE 802.11a supports far higher data-rates than 802.11b. However, it fails to support differentiated data rate services over the OFDM carrier, as a result a application implementing an 802.11a interface has to either implement the full 54 Mbps modem capability or require that during its transmission all other stations must await until the low-rate transmitter can get off the carrier.

As a result, 802.11a has not met wide acceptance by European regulators or markets, who favour the ETSI HiperLAN/2. To overcome such limitation, ETSI and IEEE have jointly formed the 5GHz Partnership Project (5GPP); this aims to merge 802.11a and HiperLAN2 into a single standard, tentatively known as the 5GHz Unified Protocol (5-UP).

By allocating a variable number of channels on a device transmission, this standard differentiates between different data rates and usage models. This is achieved by allocating the carriers within the OFDM signal on an per-device basis; multiple devices simultaneously transmit to an access point utilising different OFDM carriers. In this manner the 5-UP specification enhances the existing IEEE 802.11a standard by catering for cost-effective interface implementations supporting a wide spectrum of devices - from cordless phones to HD-TVs - over a single wireless network.

pico-cellular networks: Personal/Home Area Networks

Bluetooth A wireless technology intended to serve as a low cost, universal air interface in replacement of a large number of cable interconnects for a variety of personal devices [21]. It is supports a short-range (10m) frequency-hop wireless link, providing up to 1 Mbps in the unlicensed 2.4GHz band. Bluetooth supports both point-to-point and point-to-multipoint connections, with a maximum of seven (slave) devices communicating with a single (master) device. Its supported service discovery function allows for several pico-nets to be linked together, enabling flexible ad-hoc node configurations.

HomeRF It is an effort that aims to tackle the interoperability limitations of many 2.4GHz wireless networking products. It is supported by the HomeRF Working Group (HRF-WG), formed to establish the mass deployment of interoperable wireless networking access devices both for local and Internet data communications including voice, data and streaming media.

At the same time a growing part of the home networking industry is aligning with the HRF WG to develop the Shared Wireless Access protocol (SWAP), for radio-based home networks. The SWAP specification reconsiders interoperability through a new, common air interface that supports both wireless voice and LAN data services in the home environment with higher data rate provisions.

It is interesting to note that HomeRF/SWAP integrate DECT [367] as a dedicated transport for voice applications while IP is primarily for data streaming applications.

Appendix C

Background literature Review supplement

This annex present supplementary description of individual protocols and their signalling performance, where necessary. In addition, it provides a brief view of alternative approaches with respect to naming and routing for the purposes of mobility. Although these approaches are geared towards ad-hoc networks they aid in providing alternative insight

Section C.1 an overview of Mobile IPv4 and its signalling performance.

Section C.2 presents a brief overview of identity systems used to manage both host routing and identity.

Section C.3 presents an in-depth description of current and emerging micro-mobility proposals targeting to optimise signalling overheads and, indirectly, end-to-end signalling latency.

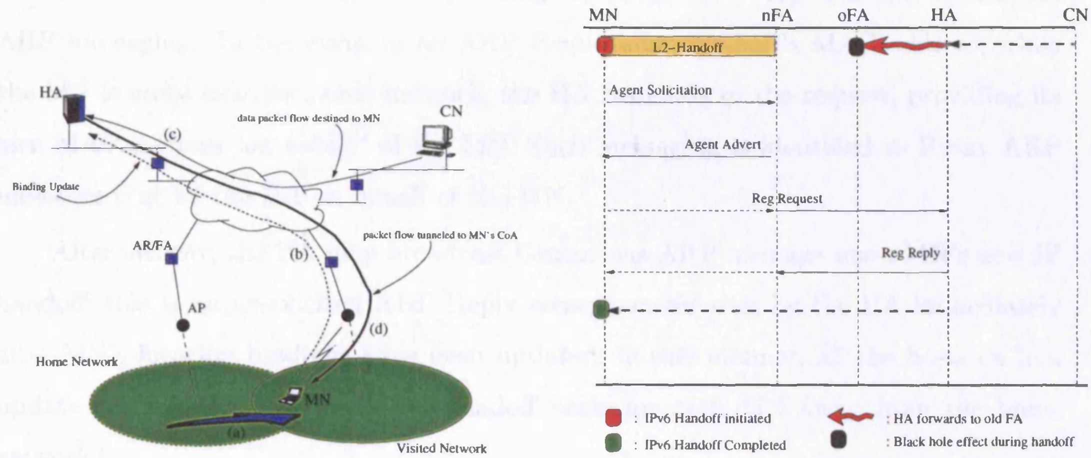
C.1 Mobile IP(v4)

Mobile IPv4 (MIPv4) sustains host reachability and routing by associating the mobile node (MN) with two addresses, while preserving the addressing and routing model of Internet: (i) its permanent home address assigned while resident at its *home* network (ii) a temporal *care-of* address (CoA) attained and refreshed at each *visited* network [368]. A visited network is identified as a subnet with a different subnet identifier and default route.

While away from its home network, MN's *permanent* home address acts as an *host* identifier for the purposes of identity resolution and security. On the contrary, its *temporally*-acquired CoA address serves as a *location*, as well as, *routing* identifier; both latter characteristics act in combination for the purposes of packet forwarding towards the MN at its new PoA, from its communicating peers.

Two principles become apparent from the above MN addressing configuration, differentiating between *IP stationarity* and *IP mobility*. The first principle dictates that while the MN remains *IP stationary*, identification/location and routing semantics overload the permanent *home* IP address associated with the MN. For the purposes of mobility management, the emerging second principle prescribes that, overloaded semantics of MN's home IP address are being *split*, to effect *permanency* in the host identification but *dynamics* in its network location and routing.

The above is achieved, by introducing a level of indirection between the *existing* permanent and a *dynamically*-acquired temporal IP address allocated to the MN. Through this norm, MIPv4 enforces a transparent mapping between the home and visited network of the MN on a per network-transition basis; this is achieved by mapping MN's home address onto the CoA acquired at the visited PoA. The mapping is instigated when the MN performs an *IP handoff* between two different IP subnets.



(a) MIPv4 handoff (b) MIPv4 signalling

Figure C.1: Mobile IPv4 operation and signalling

To assure that the MN remains *reachable* while away from the home network, the aforementioned address mapping, commonly referred to as *address binding*, is maintained by a serving host entity, present at the home network of the MN. In IP mobility management parlance, such host is identified as MN's *Home Agent* (HA). The HA is empowered with the capacity to 'defend' MN's home IP address, by acting on its behalf, while the MN is away from its home network

Figures C.1(a) and C.1(b) illustrate the modus operandi for MIPv4 and the accompanied signalling. The MN initiates its movement path from its home network making normal use of its home address, while in packet communications with the CN. Upon its departure from the home network the MN performs an IP handoff (a) once its link-layer handoff has completed. At this stage the MN solicits a Agent Advertisement so as to configure a new CoA address.

Depending on whether the CoA configured is an FA address[1] or a co-located one [2], a registration request is issued either via the FA or directly by the MN towards the HA (b). The HA creates a binding entry associating MN's home address with its current CoA.

The change of MN point of IP attachment is transparent to the CN who continues to send packets to MN's home address (c). The HA acts as an *interceptor* of MN's incoming packet traffic, independent of the application service. This is achieved by means of enforcing on-link, an address resolution of MN's address onto its own link-layer (MAC) address. To 'intercept' MN's traffic, the HA identifies itself as the MN over the home network link through two potential methods: (i) *Proxy* ARP (ii) *Gratuitous* ARP messaging. In the event of an ARP Request for the MN's MAC address, while the MN is away from its home network, the HA responds to the request, providing its own MAC address *'on behalf'* of the MN. Such messaging is identified as Proxy ARP messages sent by the HA on behalf of the MN.

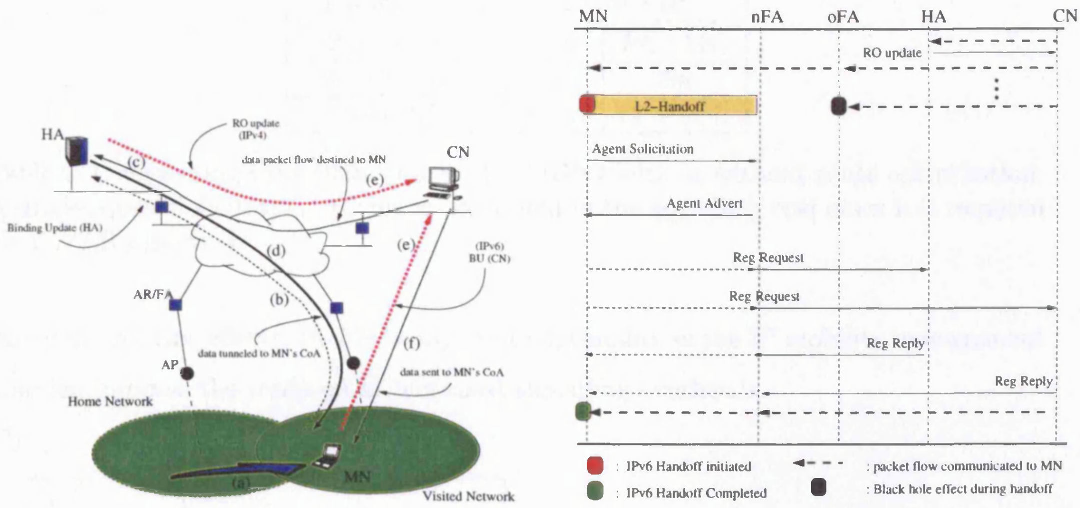
Alternatively, the HA may broadcast Gratuitous ARP message upon MN's new IP handoff; this is an unsolicited ARP Reply advertisement sent by the HA immediately after MN's location bindings have been updated; in this manner, all the hosts on-link update their ARP cache on a per-handoff basis for each MN away from the home network.

Packets received by the HA are subsequently *tunnelled* to MN's CoA (d). Depending on the type of MN's CoA, packets are either received by the FA and subsequently relayed to the MN or received directly by the MN.

It can be seen from figure C.1(b) that the MIPv4 signalling is extremely simple (2 signalling handshakes). At the same time, however, packets originating from the CN follow a considerably suboptimal route typically referred to as *triangular* routing; this is more important for *inelastic* applications such as IP telephony or video, than *elastic* types of IP traffic such as E-mail or HTTP, with respect to *latency.*

[1]the tunnel end-point terminating at the FA

[2]the tunnel end-point terminates at the co-located CoA



(a) MIPv4 handoff (with RO)

(b) MIPv4 signalling (with RO)

Figure C.2: MIPv4 operation and signalling with route Optimisation

To alleviate sub-optimality in the routing path between the MN and its CN peer(s), MIPv4 was extended by introducing the mechanism of route optimisation [369]. Figure C.2(a) elucidates the effect of route optimisation; on receipt of the first packet destined for the MN, the HA signals the CN of the MN's CoA address (e)[3]. Subsequently, the CN encapsulates any packet traffic destined to the MN and sends them to MN's CoA address (f). In this manner, the HA *redirects* the tunnel end-point from itself to the CN, while triangular routing between the MN and its peers is eliminated.

With or without route optimisation both MN and CN must implement mobility management functionality to ensure that packet headers reflect the original home address of the MN in their communications.

C.1.1 Signalling Performance

Figure C.2(b) contrasts the signalling overheads incurred by MIPv4 extended with route optimisation; the number of signalling handshakes becomes three or more depending on the number of communicating CNs.

Table C.1 presents the signalling cost[4] of MIPv4 from the perspective of the MN and the FA. It becomes apparent that route optimisation is afforded at the cost of increased signalling overhead, in comparison to its unoptimised counterpart experiencing

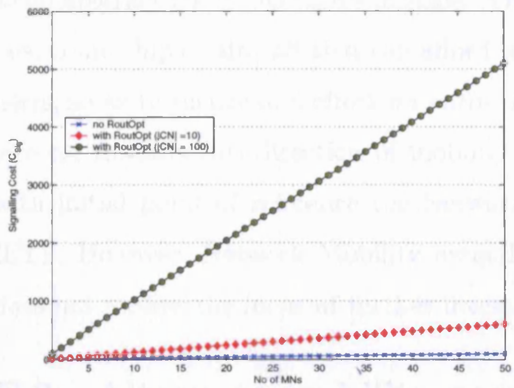
[3] depending on the protocol family, the RO is either signalled explicitly (IPv4) by the HA, or implicitly by the MN through a Binding Update (UB)

[4] The fact that in Mobile IPv6 AR discovery is required by default in IPv6 does not make the handshake a free signal. Lack of such signal can paralyse the IP mobility protocol without similar effects to IPv6 operation

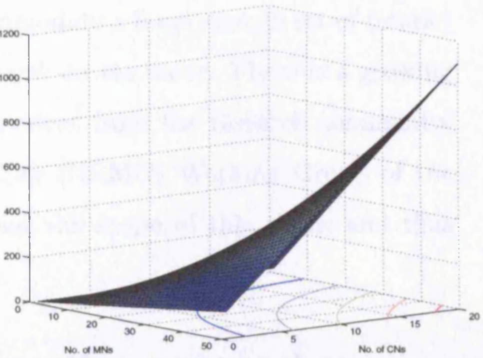
Signalling Cost	Mobile IP	
Perspective	MN	FA (AR)
without RO	2	$2m$
with RO	$2+n$	$(2+n)m$

Table C.1: Signalling Cost (handshakes) for MIPv4 with or without route optimisation. A single Agent Solicitation/Advert is accounted in the signalling cost since it is required for FA/AR discovery.

triangular routing effects; that is to say, route optimality in the IP mobility management function imposes the trade-off of increased signalling overheads.



(a) C_{Sig} of MIP with or no Route Optimisation



(b) variation of C_{Sig} in MIP with RoutOpt

Figure C.3: Comparative Signalling Cost (C_{Sig}) performance for MIPv4 with or no Route Optimisation as well as the associated variation of such cost as a function of the number of MNs each communicating with a small number of CN peers.

Figure C.3(a) presents the signalling cost of MIPv4 with or without route optimisation. It is evident that the signalling cost is significantly higher when route optimisation is employed, in contrast to the MIP signalling cost characterising triangular routing effects.

Figure C.3(b) presents a more detailed view of the variation of signalling as a continuous function of a small number of CNs communicating with each MN. The spine of the plotted surface indicates clearly that the number of CN's communicating with each MN contributes non-linearly to the buildup of signalling traffic in comparison to the number of visited MNs. From the perspective of the FA/AR this implies that small increases on the number of CNs per MN produce significantly more MIP control signalling in comparison to a large number of MNs with three or less communicating CNs.

The latter has important implications in the event that the MN acts as a router of a mobile network segment, serving multiple hosts each communicating with at least one CN. Such is the case of *network* (as opposed to host) mobility [370]; for such mobility scenarios entire network segments represented by mobile routers, manifest the abstract (super)-MN entity that multiplexes user flows from multiple attached hosts destined to multiple different correspondent nodes. In this manner, the multi-user network segment served 'hidden' behind the communicating mobile router, appears to the point of attachment as a mobile node with *multi-CN* packet flow transmissions.

Healthy examples of such mobility scenarios are met as soon as the mobile 'user', as an abstract entity, increases in scale. This is the case for vehicular entities such as a bus, train, ship or aircraft that can afford to accommodate a large enough set of (static) users, so as to justify and effect *an entire IP network on the move.* There is a growing interest towards this direction of mobility management from the research community with initial point of reference the Network Mobility (NEMO) Working Group of the IETF. However, Network Mobility extends beyond the scope of this thesis and thus does not receive the focus of further investigation.

C.2 Alternate mobility naming/routing approaches

A number of other research efforts have looked at alternative naming schemes to support addressing semantics; in such schemes the name indicates a possible set of destinations based on quantitative metrics. While such scheme have found direct application to sensor or peer to peer networks, they provide contrasting approaches to mobile host identity compared to the doctrine of standard IP mobility management.

The *Intentional Naming System* (INS) proposal [371] introduces a resource discovery protocol designed for mobile hosts. Under INS, service or resource names consist of a (key,value) tuple hierarchy. An application-level network overlay emanating from a set of self-organising resolvers, identify intended service requests; through late-binding they effect association of identified resources with routes in response to expressible interest by peers. INS relies on existing DNS mechanisms.

The *Directed Diffusion* proposal [372], transcends the notion of *intention* or *interest* by injecting an expressible availability of *interest* for certain information into the network; such interest floods the network such in turn, creates a reverse-gradient [372]. Responses follow this gradient back to the original source.

Under the *GeoCast* mobility management approach [373], packets are addressed to some location in the form of a single point or a bounded polygon region. Its routing function maintains coordinates of the total area reachable through each of its interfaces; to route traffic towards a destination, the GeoCast router calculates the intersection of that area with the GeoCast destination.

In all of these cases, changing the semantics of addressing has a direct effect on the processes of host location and packet routing. It can be seen that the techniques introduced by the aforementioned proposed alternatives, follow a contrasting different approach to the one of Mobile IPv6; while they introduce an original angle to the problem of addressing and routing in IP mobility management, they also require far more involved mechanisms with improvements, that overall, may be arguable.

C.3 Refining Macro-mobility: Micro-mobility

From the above description of dominant IP mobility standards such as Mobile IPv6, it can be seen that the underlying management function, entails essentially two core component processes:

- Handoff Management; the process by which the MN transits between successive points of IP attachment while sustaining on-going IP communications with peers
- Location management: the process which enables the mobility management function to track the IP location of the MN by means of its care-of address (CoA).

While macro-mobility approaches are characteristic of their simplicity and reduced signalling requirements, they introduce location binding updates to MN's peers on a per IP handoff basis. For location management, such requirement incurs a constant signalling overhead per mobile node. Additionally, it introduces further associated round trip time latencies with potentially adverse effects for MN's application performance. The delay incurred by location updates may be augmented if the binding update is authenticated. This is one of the fundamental trade-offs arising from simplifying the architectural paradigm of cellular networks when ported over its packet-switched Internet counterpart.

From a design perspective, the introduction of latency as a result of binding updates, attains two viewpoints: (i) round trip time (RTT) as a latency externality over which the mobility management function has little or no control, while in need of *simplistic* control signalling; (ii) the mobility management function must be conscious of the RTT variability as an external delay factor; as such it must ensure that appropriate steps are taken to reduce its emergence.

The existing Mobile IPv6 standard adopts the first design choice. While away from its home network, multiple RTT components arise between the MN and its HA as well as its communicating CN peers. However, the requirement for scalable deployment and interactive application service performance calls for attention at the second design choice. To this end, micro-mobility schemes are set to refine the performance of existing macro-mobility standards.

C.3.1 Micro-mobility principles

In essence a micro-mobility management mechanism borrows from a more refined architectural analog of cellular communication systems; it attempts to improve the performance of MN's IP handoff with the introduction of *additional* mobility agents, at the edge of a network domain. From a design standpoint, it can be seen, that system complexity is readily traded for mobility performance[5].

Micro-mobility has been primarily concerned with reducing the signalling overhead towards the home network and MN's peers. The underlying problem being addressed is that macro-mobility as currently manifested through MIPv6, requires *frequent* signalling end-to-end; that is, MIPv6 requires each MN to update its location bindings with its HA and all of its communicating CNs, on a per IP handoff basis.

To address this problem *localised IP mobility management* (IP LMM) mechanisms [374] attempt to exploit MN's *locality* of mobility reference [375]; this encompasses the property that a mobile host does not move frequently out of its provisioning domain. Hence a micro-mobility protocol attempts to mask localised IP mobility from the rest of the network by eliminating end-to-end host signalling for IP relocations *within* a domain, as shown in figure C.5(a). This design approach differentiates immediately between two types of IP handoff: (i) *intra-domain*, namely within a domain (ii) *inter-domain*, namely across domains.

By localising IP micro-mobility management mechanisms attempt to provide collectively significant reductions in signalling overheads, and latency dependencies on control signalling for the completion of MN IPv6 handoff. To this end, the architectural framework of IP LMM protocols, introduces a third IP address referred to as the *Regional care-of address* (RCoA), in addition to MN's home[6] care-of address. The RCoA

[5] justifying the design complexity met in cellular networks.

[6] as well as link-local address under IPv6.

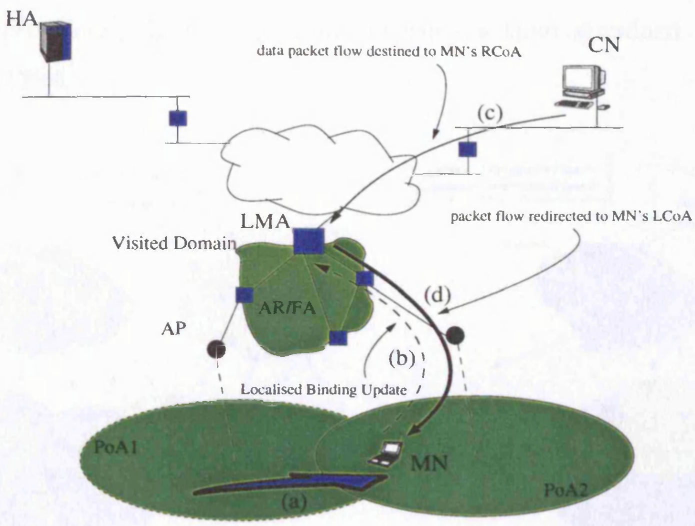


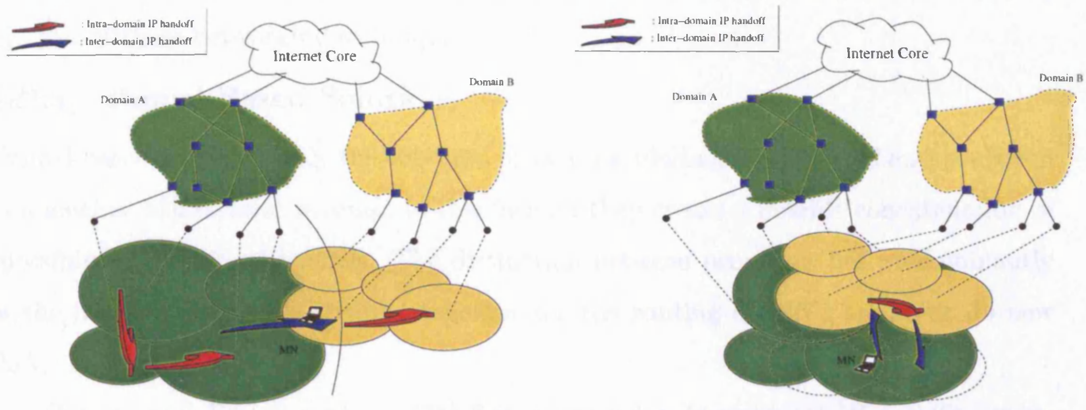
Figure C.4: Architectural framework of IP Localised (aka micro-) Mobility Management (IP LMM) protocols

address represents MN's globally reachable care-of address for the purposes of traffic forwarding and location updates at both the HA and the CNs. Intra-domain IP mobility is handled through localised location binding updates to a *Local Mobility Agent* (LMA) [374]. Incoming packets are routed to the LMA, as they are destined to MN's RCoA; on arrival at the LMA, the latter *'redirects'* them to MN's *local* CoA (LCoA). Figure C.4 illustrates the architectural framework of IP LMM protocols.

Signalling savings from localised routing decisions are significant as long as high mobility is effected by the MN *within*, a single provisioning domain. This is not necessarily the case when high mobility is effected across multiple access provisioning domains [46].

We have shown [44] that, in fact, application of IP LMM protocols in cases of frequent *inter-domain* IP mobility has adverse effects on signalling overheads, compared to Mobile IPv6 signalling overheads; that is to say, micro-mobility protocols incur *higher* signalling overheads when the rate of inter-domain IP handoffs increases significantly in comparison to the rate of intra-domain handoffs.

This above is the case in when the MN increases its handoff rate in *vertical* IP handoffs *within* a network domain as shown in figure C.5(b). The MN *must* update its binding *both* at intra-domain and inter-domain level. IP handoff oscillations between different network domain as a result of capability, tariff or QoS dynamics, generate an adverse effect on signalling overheads that is *counteracting* the fundamental requirement of micro-mobility protocol; this is because frequent vertical handoffs, micro-mobility mechanisms incur clearly higher signalling overheads than standard macro-mobility, namely Mobile IPv6.



(a) Horizontal IP Handoffs (b) Vertical IP Handoffs

Figure C.5: Increase in inter-domain handoffs as a result of multiple vertical handoffs over different network domains within a certain geographical area

As we see in subsequent sections, lack of signalling optimisations for vertical handoffs is the fundamental limitation in *all* micro-mobility protocols. This is introducing limited scalability in signalling performance over future ubiquitous multi-wireless network-overlay infrastructures envisaged as the next generation IP wireless-access environments.

Furthermore, what is interesting to note in terms of latency is that, while micro-mobility protocols *do* manage to localise signalling latencies, they still remain under the influence of increased RTT variability within a network domain, as discussed in Section 2.3.3. Such variability accounts for 75% of the regional paths experiencing RTT variations above 50ms and up to 200ms.

C.3.2 Current and emerging IP Micro-mobility management solutions

Despite adverse performance over vertical handoffs, micro-mobility protocols can reduce significantly the number of location updates as a result of high mobility within a single network domain.

Depending on the design assumptions and techniques adopted, different micro-mobility proposals present different scaling properties. To this end, they consider different design trade-offs and experience different performance characteristics.

From this perspective, micro-mobility mechanisms may be classified as a taxonomy of three broad *redirection* techniques: (i) *tunnel-based* (ii) *routing-based* and (iii) *multicast-based* IP LMM protocols.

Based on this taxonomy, the following sections survey the most important advances on micro-mobility management; we outline issues and challenges addressed or emerging by each of these networking techniques.

C.3.3 Tunnel-Based Solutions

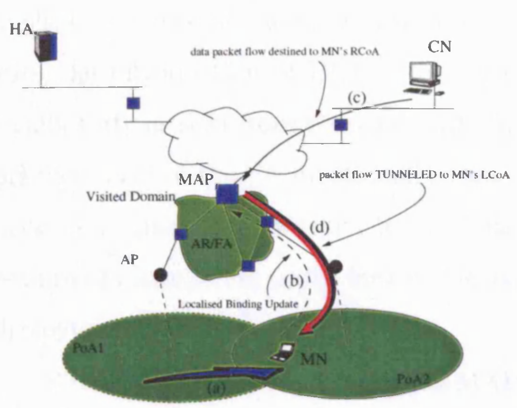
Tunnel-based schemes apply the concepts of location binding updates and encapsulation in a local or hierarchical manner; in this fashion they create a flexible concatenation of (possibly several) local tunnels. The distinction between proposals lies predominantly in the number of tunnels chained together for the routing of MN's traffic at its new PoA.

Caceres and Phadmanaban [376] first proposed a *hierarchical* IP mobility management scheme based on Mobile IPv4. Their proposal defines a *two-level* hierarchy of FAs. In this proposal, each subnet visited by the MN has one (or more) FAs. At the root of the FA hierarchy, a domain FA manages mobility across subnets within the network domain. MN's HA keeps track of MN's movement across administrative domain boundaries. As a result, MN's motion within an administrative domain remains transparent to its HA and communicating peers. Gustafsson [377] expands on the HMIP architecture of the Caceres et al with a similar architecture using a *multi-level* hierarchy of FAs identified as MIPv4 with Regional Registrations (MIPv4Reg). In this scheme FAs identify the *crossover* point between FAs and as a result increases the degree of localisation in both signalling and its associated latency.

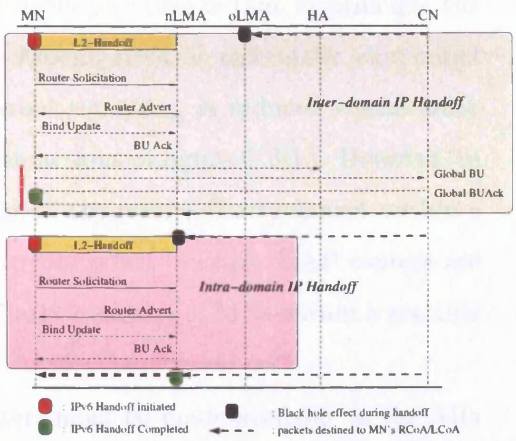
While lower signalling overheads and latency can be observed in both of these proposals, their basic limitation stems from the requirement of FAs in every point of attachment. In addition, the MIPv4Reg proposal requires further the introduction of FA within the core of the administrative domain, so as to formulate its multi-level routing hierarchy. This implies effectively the introduction of host routes within each FA and the bypassing of dynamic routing mechanisms, in view of crossover-based routing.

The Caceres et al. proposal is subsequently optimised with extensions over of IPv6 proposed by Castellucia [378]. This is augmented by the HMIPv6 protocol specification proposed jointly by Soliman and Castellucia et al. [30] in the IETF. The proposal encourages the perspective of localisation of the HA in the locality of the domain visited by the MN; the former is emerging as the root local mobility agent (LMA) also referred to as Mobility Anchor Point (MAP). This is shown in figure C.6(a).

Table C.2 provides the measure of signalling cost for HMIPv6 versus the respective



(a) HMIPv6 handoff



(b) HMIPv6 signalling

Figure C.6: HMIPv6 operation and signalling

Signalling Cost	MIPv6 with RO		HMIPv6	
Perspective	MN	AR	MN	AR
Intra-domain	$2+n$	$(2+n)m$	2	$2m$
Inter-domain	$2+n$	$(2+n)m$	$3+n$	$(3+n)m$

Table C.2: Signalling Cost (handshakes) for MIPv6 with RO and HMIPv6. A single Solicited Router Advertisement is included in the signalling cost since it is essential in MAP discovery

cost of MIPv6. Clearly, both the MN and AR experience higher signalling overheads over HMIPv6 for an increasing rate of inter-domain handoffs.

MIPv6 Regional Registrations In a rationale similar to MIPv4Reg, Perkins and Malinen propose a similar form of hierarchical mobility, termed as MIPv6 Regional Registrations (RegRegv6) [230].

In a fashion similar to MIPv4, RegRegv6 attempts to approach also some form of signalling distribution by employing the notion of crossover LMM agent router, such that regional signalling need not propagate all the way up to the gateway LMM agent. This, however, does not reduce or load balance the gateway LMM agent, and as such it still suffers from scalability issues; scalability fails also to be addressed through auto-configuration and placement of the LMM agent in optimal paths that avoid further introduction of triangular routing, with respect to MN's handoff path.

Much like its IPv4 predecessor, RegRegv6 does introduce multiple levels of routing hierarchy with host routes which counteracts the dynamic routing function of existing routing engines, and subsequently amplifies its scalability limitations [46].

HMIPv6 The improvement brought by the HMIPv6 proposal is that in principle the explicit existence of the FA in the form met in Mobile IPv4, is essentially eliminated with the introduction of IPv6. In addition global signalling is reduced significantly significantly as seen from the set of signalling interactions of figure C.6(b). However, in practice, part of the FA functionality *can* emerge at the points of attachment within a network domain in an *implicit* form. This is the case when *multiple* MAP routers are required to handle the traffic forwarding load of large numbers of MNs within a scalable deployment scenario[7].

Since each RCoA of the multiple MAP routers must be made available to the ARs to advertise it to the MN, the HMIPv6 proposal encounters significant complexities of MAP configuration, particularly for large network domains. In such cases the emerging problem is how will the RCoA addresses of multiple MAPs be configured at ARs such that: (i) each MAP can be either load-balanced or support replication consistency to avoid single points of failure (ii) each segment of the network domain is serviced by *exactly* one MAP (iii) the correct MAP at the edge of the network domain will service the MN originally associated with [46].

To this end, HMIPv6 identifies the outline of a MAP discovery mechanism which, however, requires propagation of MAP options towards the ARs; that is diffusion of multiple MAP discovery signals across the entire domain [301]. It can be seen that for multiple MAPs such discovery mechanism introduces more signalling within the domain, than the original signalling overheads incurred by macro-mobility mechanisms. We remind that the amount of such signalling becomes a fixed overhead signalling cost for HMIPv6 since MAP can dynamically change their preference options for load sharing purposes. We argue that such signalling and configuration complexities counteract the benefit of localised IP mobility management

On the contrary, in the case of a *single* MAP all ARs can be configured with a single RCoA. However, this introduces significant concerns over both scalability and reliability over single points of failure, since it concentrates the traffic load of the entire set of accommodated MNs through a single mobility management gateway [46].

HMIPv6 [301] attempts to approach auto-configuration through extensions in the router renumbering process of the core IPv6 protocols. However, there is no evaluation for the proposed auto-configuration approach with respect to the rate of convergence on LMM info availability to the MN, especially when a *new* LMM agent is introduced

[7]Such issue arises also in the case of MAP replication aiming to fight single points of failure

in a domain.

In latest revision of its draft specification HMIPv6 allows the MN to register with multiple MAPs (if available) for the purposes of efficient usage of access network bandwidth. This however, creates significant replication and redundancy of regional LCoA registrations to multiple MAPs, augmenting the total amount of intra-domain signalling for large number of MNs. It can be seen that such technique re-introduces the problem of significant signalling overheads, this time, localised within the domain.

It is noted that none of the aforementioned approaches deals with issues of state establishment, in particular IPv6 addressing. This implies that existing tunnel-based solutions fail to preserve IP mobility seamless from a delay transparency perspective.

It is worth noting S-MIP [379] as a derivative of HMIPv6 that aims to enhance the handoff process by means of movement pattern heuristics; it employs signal strength measurement obtained from the link-layer of APs, in an attempt to predict an impending IP handoff by the MN. The S-MIP proposal presents improved handoff performance by means of introducing an additional mobility management entity identified as decision engine (DE). A DE has the same scope as a MAP agent by maintaining an intra-domain view of the mobility pattern of all MNs transiting within that domain. It can be seen that such proposal introduces additional scalability limitations by concentrating the prediction processing and signalling at the MAP. It however, suggests that under particular movement scenarios predictive techniques can achieve significant improvements in IP handoff delay performance.

C.3.4 Routing-Based Solutions

For routing-based solutions, the micro-mobility management function, exploits solely the robustness of conventional but simple IP forwarding. This is complemented by introducing special routing functionality at network layer, serving the specification of the particular mobility management design.

Borrowing from cellular architectures, routing-based micro-mobility schemes introduce a mobile host location database that is created and maintained within a domain in a distributed manner [130, 380]. The database consists of individual flat mobile-specific identifier[8] lookup routing tables, while maintained by all mobility agents within the serving network domain.

Such techniques are employed by HAWAII [229] and Cellular IP [130] protocols with distinct approaches only in the functionality of the nodes and the construction

[8] address or address-like

methods of the routing tables. The following sections presents briefly its operations together with associated benefits or limitations arising from their deployment.

HAWAII

Ramjee et al. proposes in [229, 381] the HAWAII micro-mobility management architecture in support of efficient intra-domain mobility management. In this proposal, a network domain comprises of IP routers supporting HAWAII-specific host routing. To allow this the HAWAII architecture requires the existence of a domain gateway, handling all incoming or outgoing MN traffic; such gateway is identified as Domain Root router (DRR).

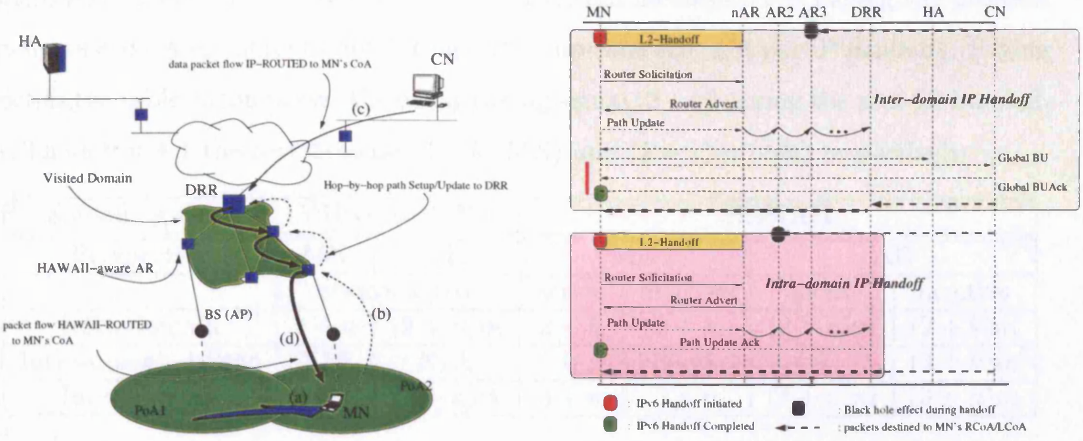
It is important to distinguish between a mobility router and a mobility agent; a DRR is a routing engine, *not* a mobility agent. A mobility agent (e.g HA) may exist independent of the routing device although its performance improves if both are co-located. A mobility agent typically *intercepts*, *tunnels* and *redirects* traffic for the MN towards MN new point of attachment (PoA); a mobility router *routes* traffic destined to the MN with no need either for interception (ARP resolution) or tunnelling.

A DRR router exists at the edge of any network domain, whether home or visited. Traffic destined for an MN at the home domain is routed always through the DDR towards the MN.

HAWAII adopts the notion of *Home (network) Domain*, as opposed to notion of *home subnetwork* typically employed in tunnel-based micro-mobility solutions (e.g HMIPv6). For *home* intra-domain movement of the MN between successive PoA, the DRR routes *always* traffic towards the MN with no intervention by the HA; that is, the HA *does not* intervene when the MN changes its point of IP attachment *within* its home domain. This is shown in figure C.7(a)

When the MN transits onto a foreign network domain, the HA intercepts the traffic routed to it by the DRR and tunnels it to MN's new CoA. The immediate limitation arising from this is that unless the DRR and the HA are co-located such forwarding approach results into significantly sub-optimal routes. At the same time co-location imposes significant traffic load bottlenecks as well as configuration complexities in the event of multi DRR configurations.

The DRR routes the packets towards the MN using host-routes stored at each HAWAII-aware router, on each hop along the path. When the MN moves between different subnets of the same domain, only the intra-domain route between the DRR and the serving base station (BS) is modified. Hence, in a fashion similar to tunnel-



(a) HAWAII handoff (b) HAWAII signalling

Figure C.7: HAWAII and CIP operation and signalling. It should be noted that in CIP the signalling is conditioned by the activity of the MN. The above reflects the case when the MN is inactive with no paging enabled

based solutions (such as HMIPv6), global signalling load and handoff latency, as a result of inter-domain RTT variability, can be reduced.

To establish and maintain a dynamic path to the MN, HAWAII uses *explicit* path setup messages, as shown in figure C.7(b). These messages establish a host-specific path from the DRR to the MN by creating host-specific forwarding entries in the routers along the path, commonly referred to as *host routes.*

Whether a power-up or an inter-domain IP handoff, the MN is first assigned dynamically an IP address. Within the domain such address is used *solely* for identification purposes; this is because routing is handled explicitly by the path enabled through host-route setup. Outside the domain the address maintains it standard IP semantics identification and addressing, until the packet arrives at the DRR.

Path setup state (host-routes) maintains a lifetime (soft-state); the MN must refresh such state by means of periodic signals sent to its current AR. The AR subsequently propagates this message to the next hop along the path to the DRR. In this manner, the micro-mobility protocol *maintains* an active intra-domain route between the DRR and the MN. Table C.3 presents the signalling cost of HAWAII in terms of Path setup signalling overhead from the perspective of both the AR and the MN, compared to MIPv6; n represents the number of CNs, m represents the number of MN's and k the number of path refresh messages per IP handoff.

The typical signalling cost encompasses a handshake for address allocation and a handshake to inform the DRR and any path refresh messages. For paging the protocol incurs one extra signal sent only during call-setup time (i.e. not per IP handoff). Paging cost in the table encompasses the extra paging signal $(3+k)$ during the n-th IP handoff. At handoff $n+1$ the cost becomes $2+k$ (MN) and $(2+k)m$ (AR) respectively.

Signalling Cost	MIPv6 with RO		HAWAII			
Perspective	MN	AR	MN		AR	
	always active		active	inactive	active	inactive
Intra-domain	$2+n$	$(2+n)m$	$2+k$	$2+k$	$(2+k)m$	$(2+k)m$
Intra-domain+Page	N/A	N/A	0	$3+k$	0	$(3+k)m$
Inter-domain	$2+n$	$(2+n)m$	$3+n$	$3+n$	$(3+n)m$	$(3+n)m$

Table C.3: Signalling Cost (handshakes) for MIPv6 and HAWAII. A single Solicited Router Advertisement is included in the signalling cost since it is essential in DRR discovery

In addition, the DRR maintains a flat address lookup table with forwarding metrics for *all* active MN within its domain, while each routing node maintain part of this table.

Furthermore, HAWAII makes forwarding provisions during the period of the IP handoff in four different approaches:

- multi-stream (MSF) : by means of propagating path update messages from the new BS, each hop emanating from the old AP, forwards MN's traffic to the new AP where the MN is.

- single-stream (SSF) : similar to MSF with the difference that only the old AP can forward packet to the new AP.

- unicast non-forwarding (UNF) : the crossover router unicasts incoming traffic to the new AP

- dual-cast non-forwarding (MNF) : the crossover router *dual-casts*[9] incoming traffic to both new and old APs

According to the authors of the HAWAII protocol, UNF and MNF at the crossover router exhibit the best performance, with MNF consuming marginally larger buffer size at the routers.

For idle MNs, HAWAII supports also an IP paging function. It uses IP multicast to page idle MNs when packets destined to an MN arrive at the DRR while no recent routing information is available.

[9]commonly known as bi-casting

Limitations As it may be seen from the signalling requirements of the HAWAII protocol, in routing-based micro-mobility management the location update signal is effectively replaced by the path update message propagated hop by hop. In addition, the scheme requires the introduction additional routing state which grows linearly as a function of the number of MNs. In effect, such design approach requires changes not only at the access routers but in all routers within the domain, with scalability concerns about the growth of host-route state under real deployment scenarios for large number of MNs.

Additionally, HAWAII remains an intra-domain protocol. Inter-domain movement is handled by MIPv6. For frequent inter-domain handoffs the scheme, introduces additional signalling overheads while lacking delay transparency as a result of IP state establishment.

What's more, while Hawaii can incur significant signalling savings by confining location management to intra-domain routing signalling, it suffers considerably from single points of failure; only the routers along the path between the serving AR and the DRR know *how* to route packets towards the MN over the access network (intra-domain) [382, 383]. If either the DRR or any of the routers along the path fails then the forward path towards the MN breaks and the mechanism fails to deliver traffic to the MN. On the contrary, by means of dynamic routing all routers know how to route packets towards the MN.

Ultimately, while the soft-state character of signalling is considerably appealing, it effectively inflates the signalling cost of the micro-mobility protocol. In fact, it appears that if the number of path refresh messages is bigger than the number of MN communicating peers[10], HAWAII introduces higher signalling overheads than MIPv6 or HMIPv6.

Cellular IP

Cellular-IP (CIP) proposed by Valko et al. [228, 130] is another important routing-based micro-mobility management mechanism. CIP shares a number of similarities with HAWAII while it adopts a different signalling and routing approach.

CIP identifies for mobility management purposes its own means of routing within a network domain. Inter-domain mobility is handled by MIPv6 and routing by IPv6, while both intra-domain mobility and routing is handled by CIP. Individual network domains are connected to the network backbone by means of routing *gateways* (GW).

[10] it can be seen that this in fact may be the common case: while the MN may communicate on average 1-2 CNs at any time, we anticipate that path updates will be greater on average than 1-2 refresh signals along the path to the DRR.

Similar to HAWAII, the forwarding lookup table of GW contains entries for all active MN in the domain. Outside a network domain, any MN is identified by the IP address of that CIP gateway

Instead of path setup messages explicitly sent by HAWAII, CIP instigates on routing nodes the ability to *learn* the source IP address of upstream data packets and derive from them the corresponding downstream interfaces. The upstream path towards the domain CIP-specific router is inferred by each AR within the domain using beacon packets periodically transmitted by the GW; all packets sent upstream by MNs are routed towards the gateway using this path.

Perhaps the greatest contribution of CIP is the differentiation, with respect to signalling, between active and idle MN in its location management function. While active MNs update their downstream route by *on-going (data) packet transmissions*, idle MNs send *periodically* explicit *route update* messages towards the GW. If active during a handoff, the MN updates its downstream routing path by sending its first packet towards the GW using the beacon-formulated upstream route. If inactive, the MN simply sends a route-update towards the GW using the same upstream route. Table C.4 presents the signalling cost from the perspective of MN (active or inactive) and AR.

Signalling Cost	MIPv6 with RO		CIP			
Perspective	MN	AR	MN		AR	
	always active		active	inactive	active	inactive
Intra-domain	$2+n$	$(2+n)m$	0	$1+k$	0	$(1+k)m$
Intra-domain+Page	N/A	N/A	0	C_P+1	0	$(C_P+1)m$
Inter-domain	$2+n$	$(2+n)m$	$3+n$	$3+n$	$(3+n)m$	$(3+n)m$

Table C.4: Signalling Cost (handshakes) for MIPv6 and CIP. A single Solicited Router Advertisement is *not* included in the signalling cost since CIP does not allocate addresses within a CIP network domain

By enforcing such differentiation, while exploiting on-going data packets communicated upstream to the CN to piggyback signalling, CIP achieves reductions of *explicit* intra-domain signalling. However, such reduction comes at the cost of placing *active routing* [384] processing rules on the routing engine; this implies that a CIP active router must *constantly* process the content of data packets to infer mobility signalling. Reason for this is the fact that routers do not have explicit knowledge of *when* the CIP handoff of a single MN occurs. Hence they must check the packet routed constantly to infer the potentially new crossover point emerging as a result of a CIP handoff. Such routing extensions introduce prohibitive processing load on routers of CIP network domains handling very large numbers of MNs.

Two CIP handoff schemes are supported: *hard* CIP handoff allowing some packet loss while being efficient in the amount of signalling overhead and latency; *'semi-soft'* CIP handoff aims to minimise the transient packet loss. It is noted that the use of semi-soft or hard handoff is dependent on the wireless technology used by the MN and not on the CIP protocol. To expedite handoffs, CIP exploits the crossover between CIP-routing paths, such that packets arriving downstream are instantly routed towards the new path *at the crossover point* between old and new CIP routes. This is possible under CIP routing, since the problem of maintaining a mapping between two addresses (home and CoA) is eliminated; the MN maintains its home IP address throughout its mobility pattern, whether inter-domain or intra-domain, solely for identification purposes.

CIP encounters similar limitations to the ones of HAWAII with perhaps the most significant one being the requirement of replacing the existing IP routing function; intermediate routers are required to maintain host-specific routes that are updated indirectly by packets sent by the MN. Although applicable to small, campus-sized environments, this technique meets significant scalability concerns in large-scale Internet deployment. In addition, Eltahir and Dunlop identify that CIP may not suit all network topologies since certain types introduce a significant number of crossover cache updates [385].

Furthermore, Castellanos et al [386, 382] identifies that in the event that the stream duplicate flowing towards the new AR has a shorter RTT then there is the potential of packet loss at the new access point; this can impede seamlessness for delay sensitive applications.

TIMIP TIMIP [387] is a recent combination of design principles of CIP and HAWAII. Here the IP layer is coupled with link layer handoff mechanisms of the underlying wireless technology. The fundamental difference from the above is the inclusion of a context transfer mechanism in support of expediting IP handoffs. While the solution is limited in a manner identical to CIP and HAWAII it promotes use of context transfer mechanisms similar to those currently under discussion within the IETF Seamless mobility (SEAMOBY) working group [233].

EMA

EMA [388, 389, 390] employs a different routing mechanism than standard IP forwarding/routing. This is adhoc routing, in particular, the temporally ordered routing algorithm (TORA) [391] protocol. This requires that all networks must implement the TORA features for the proposal to work. This is potentially unrealistic, since TORA has only been used in ad hoc mobile environment, not fixed infrastructure one. It is highly unlikely that routing protocols would encompass TORA as a means of routing as opposed to classic routing protocols like ISIS [392] and OSPF [393]. While the proposal is argued to be scalable [382], it remains to be seen that it is also fault tolerant, since its reliability depends on the TORA forwarding features which are not commonly deployable in infrastructure networks.

C.3.5 Multicast-Based Solutions

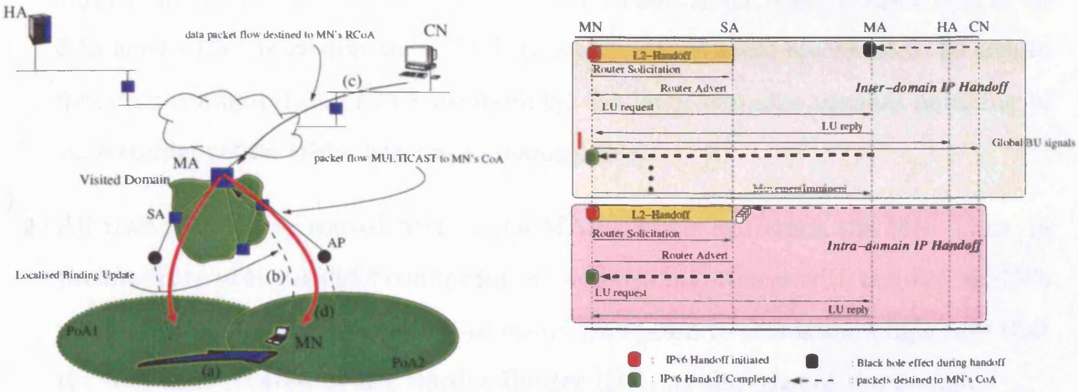
IDMP and M&M

Intra-Domain Mobility Protocol (IDMP) [231] originating from the TeleMIP proposal of Das et al [394] and Multicast Micro-mobility (M&M) [395] are two other micro-mobility protocols that localise mobility management signalling. Both schemes depart from the CIP and HAWAII designs by forwarding traffic during an IP handoff through a *multicast* group rooted at the *Mobility Agent* (MA); this is illustrated at figure C.8(a). In a similar fashion to tunnel or router-based micro-mobility protocols, the MA requires placement at the edge of the network and in particular co-located with the border router (BR) of that network domain. Since the two protocols are fundamentally similar[11] in their forwarding function, we focus our elaboration on IDMP; performance issues or benefits from the use of IDMP are identical for M&M as they stem from the use of multicast routed at the edge of the network domain, not from the design internals of each proposal.

In comparison to CIP and HAWAII, IDMP simplifies the forwarding function while capturing more robustly transient ping-pong effects, as a result of the unpredictability of MN's movement pattern relative to its point of attachment (PoA). It achieves a similar signalling budget to HAWAII for its paging function as shown in figure C.8(b); this is lower than paging signals in CIP which requires legacy paging mechanisms, since IDMP-paging is performed over a multicast group; such multicast group clusters sets of ARs so as to create bounded paging areas. However, in an manner identical to previous tunnel- or routing-based micro-mobility proposal, the IDMP approach suffers the standard LMM limitations [46, 45] described already in previous sections.

In addition, IDMP protocol introduces sub-optimal routes occurring from forwarding traffic to a fixed Rendezvous Point [185]; the IDMP proposal makes no provisions for optimisation of such issue, other than to assume that the RP is co-located at the

[11] M&M differs from IDMP in that it introduces soft-state in multicast group management and allocates a multicast address for each MN. However, the multicast decapsulation is always performed by the last hop AR in both schemes.



(a) IDMP Handoff (b) IDMP signalling

Figure C.8: IDMP operation and signalling requirements

MA. Such approach, however, amplifies the dependency on particular topological assumptions which incur both performance bottlenecks and candidate failure points. The latter has detrimental effects for operation of all MNs within a domain.

Furthermore, the period of packet loss from the moment of the handoff initiation signal sent from the MN until the moment of its receipt by the MA, is a function of the one-way delay between the MN and the MA. The latter is typically dependent on the size of the administrative domain assuming each administrative domain is serviced by one MA. However, as seen in section 2.3.3 even intra-domain the RTT variability can reach from a lower bound of 50ms as high as 200ms for 75% of the regional traffic. It implies that the RTT between the MA and AR points of attachment introduce significant delay when *signalling the initiation of a handoff.* In addition, all transient packets en route to the previous PoA, while an IP handoff has been signalled by the MN to MA, *are guarranteed to be lost.* This is confirmed also by reported results in [396], showing that IDMP does not eliminate the delay incurred by the actual handoff process[12] and requires additional protocol mechanisms.

To fight single points of failure a micro-mobility protocol is required to support multiple MAs. Such requirement introduces typically additional signalling and configuration complexities. With reference to IDMP configuration of multiple MAs per administrative domain imposes an additional configuration, route optimality as well as synchronisation/reliability task; this is because:

1. The address of each individual MA must be configured to be advertised in different

[12] particularly when extended with more context signalling such as QoS [397]

subnet clusters of the domain. This imposes an additional configuration task in *all* SAs across the entire domain. IDMP proposes no protocol mechanisms to ensure dynamic configuration of MA availability; for large domains manual handling of such configuration tasks become unmanageable.

2. All traffic has to be routed first to the MA before it can reach the MN. Thus, in the majority of topological configuration, sub-optimal routes with respect to MN's point of attachment are expected to occur; exception to this is the single case that the MA is co-located at the Border Router (BR) at the edge of the domain.

3. In the event that the MA is co-located with the BR, the scheme limits its capability of recovery from MA failure, since alternative MAs cannot be utilised (even at the cost of a suboptimal route). In the event of multiple MAs within the domain, complex consistency mechanisms are required to handle the failure of an MA by remapping MNs to a new MA within the domain.

Critical point to the operation of the multicast forwarding is the lack of reliability in the handoff initiation signal sent by the MN to the MA; whether emanating from the SA or the MN, the IDMP proposal prescribes no reliability in the transmission of such signal by the MN, in the face of increased access network congestion or loss. We remind that under the current mode of operation in IDMP there exists no guarantee that the MovementImminent message will be received by the MA before the MN detaches from the wireless link of its current PoA.

The lack of robust handoff initiation signalling is augmented by the fact that the MA is *multiple-hops away* from the PoA of the MN; hence the probability of loss for the handoff trigger signal becomes the sum of the loss probabilities on *each* individual hop towards the MA. Thus, while the trigger may work with reasonable statistical reliability on lightly loaded networks, this is not necessarily the case under heavy load conditions. It is noted that handoff initiation is critical factor for the operation and performance of any mobility management protocol mechanism at hand. It is thus of question if the IDMP approach can effect reliably the initiation of a handoff over a loaded or congested network infrastructure.

Furthermore, the IDMP proposal furthermore has no means of identifying the set of neighbour *subnet agents* (SAs) such that they can all join the same multicast group for the purposes of forwarding or paging. It is important to note that in a spanning tree network topology there is no guarantee that the neighbouring SA belong to the *same*

Signalling Cost	MIPv6 with RO		IDMP			
Perspective	MN	AR	MN		AR	
	always active		active	inactive	active	inactive
Intra-domain	$2+n$	$(2+n)m$	2	1	2m	m
Intra-domain+Page	N/A	N/A	2	2	$2m$	$2m$
Inter-domain	$2+n$	$(2+n)m$	$3+n$	$3+n$	$(3+n)m$	$(3+n)m$

Table C.5: Signalling Cost (handshakes) for MIPv6 and IDMP. A single Solicited Router Advertisement is included in the signalling cost since it is required for the purposes of address allocation within a domain

administrative domain; from this perspective it becomes increasing difficult to configure SA neighbours by manual means of topological assumptions.

It can be seen that the IDMP proposal does not eliminate the issue of handoff latency incurred by the acquisition of a GCoA (at inter-domain) or the LCoA (at intra-domain) level. This is the case for addresses that may be acquired either *statelessly* or *statefully*, that is through DHCP mechanisms (see section 2.3.4).

The delay introduced by stateful DHCP address allocation under IDMP, counteracts the amount of signalling expended for fast handoff management. This is because both inter-domain and intra-domain IP address allocation approaches incur multi-second delays (1-5sec) [398]. Such delay during address configuration, irrespective of the amount of buffering effected at the new PoA(s), is *guaranteed* to render the buffered packets as lost for the purposes of interactive communication services; any delay above 200ms renders *by definition* the buffered packets as 'lost' since they will arrive at the MN much later than their scheduled play-out deadline.

From a configuration perspective, the management of SAs under a multicast group assumes a fixed network topology. This meets fundamental limitations in cases of incremental deployment of the IDMP mechanism in routing infrastructures; the new AR neighbours have no means of attaining information that will allow them to attach onto some multicast forwarding address that describes the group of neighbouring SAs.

Ultimately, the IDMP proposal relies on standard MIPv6 for inter-domain handoffs and hence introduces the standard MIPv6 delay for inter-domain handoffs. In large wireless administrative domains, inter-domain handoff may be assumed to be infrequent if not rare for the mobility pattern of the majority of MN; as a result the signalling cost of IDMP may be deemed acceptable for *horizontal* handoffs; table C.5 provides a breakdown of intra-domain and inter-domain signalling for IDMP.

The same however, does *not* apply for *vertical* handoffs; the signalling cost grows faster at the AR as a result of multiple (vertical) inter-domain handoffs. Multiple vertical handoffs in future heterogeneous wireless networks [147] are expected to be the result of dynamic tariff, bandwidth or congestion differentiation amongst competing wireless ISP domains [224].

From the above it can be seen that the IDMP solution overall evolves satisfactorily the micro-mobility management function towards fast handoff management, but remains severely limited in terms of scalability in a wireless Internet supporting both types of (horizontal or vertical) IPv6 handoffs.

Appendix D

Experimental MIPv6 evaluation supplement

This annex provides supplementary results relevant to the experimental evaluation of MIPv6 handoff performance.

Section D.1 presents a critical view of the MIPv6 handoff process emerging from the underlying protocol design of both MIPv6 and core IPv6 protocols.

Section D.2 presents a delay analysis of key component functions comprising the MIPv6 handoff process.

Section D.3 presents the mechanism for calculating the delay variance during a MIPv6 handoff.

Section D.5.1 presents an in-depth analysis of Neighbour Discovery performance monitored over both v2h and h2v MIPv6 handoffs.

Section D.4 presents the delay distribution induced by neighbour reachability during a MIPv6 handoff.

Section D.5 presents the statistical distribution of the measure of handoff delay experienced during a v2h MIPv6 handoff.

Section D.6 presents a detailed treatise on the influence of the wireless medium, with focus on 802.11b Wireless LAN, over the MIPv6 handoff process. By means of simulations it demonstrates the effect of wireless MAC contention onto the efficiency of router advertisement interval.

Section D.7 presents the derived statistical distributions of the L2-handoff delay component arising over 802.11 WLANs, contributing to the total measure of MIPv6 handoff delay.

D.1 A detailed view of the MIPv6 handoff process

D.1.1 Movement Detection

In Mobile IPv6, it is generally the responsibility of the MN to detect that it has moved between networks. Determining whether or not a MN has moved networks is not always a simple issue. However, the general rule of thumb that a MN has moved can be seen as:

- the current access router is no longer reachable
- a new (different) Access Router (AR) is available

To determine if its current AR is still bi-directionally reachable the MN performs Neighbour Unreachability Detection on a continuous basis.

Neighbour Unreachability Detection (NUD) works in the following manner: when an IPv6 host has a packet to send, it checks the Neighbour Cache to determine the link layer address of the next hop node (either an on-link neighbour or a router). The Neighbour cache has also reachability state associated with each neighbour entry. A neighbour cache entry in REACHABLE state, indicates that the neighbour is considered reachable on-link.

In IPv6 a host considers a neighbour reachable if it has recently received confirmation that packets sent to the neighbour have been received. This is achieved in two ways: the receipt of a neighbour advertisement from the neighbour in response to a neighbour solicitation sent by the host, or a hint from upper layer protocols. The IPv6 stack utilises the acknowledgements of upper layer protocols to register the fact that a packet has recently been received from a given destination address and so is considered reachable.

The IPv6 host will send a neighbour solicitation in the event that the neighbour cache entry is not set to REACHABLE when there is a packet to send[1].

Note that the NUD function occurs only when the MN has a packet to send. Thus, for the worst-case scenario where the MN is not sending any packets, it may not notice that it has moved networks until it receives an unsolicited router advertisement from the new on-link router (consistent with the normal router advertisement interval). Unfortunately, this may be the case when the MN is receiving real-time streams when an interruption in connectivity can cause packet losses and unacceptable latency while the

[1] This may involve a wait of DELAY_FIRST PROBE_TIME seconds if the neighbour cache entry is in the DELAY state

new handoff is taking place. In such a scenario, the MN may not actually be transmitting much data itself, perhaps occasional TCP or application layer acknowledgements, but nothing that will allow the unreachability of its current AR to be discovered in a timely fashion.

It should be noted that the availability of a new router advertisement serves *only* as a hint that the MN has moved networks; it does not guarantee the occurrence of an IP handoff. This is the case where a new (additional) router has been activated on the existing link. Furthermore, as prescribed in [107] unsolicited router advertisements must *not* be used as confirmation of bi-directional router reachability since they only confirm reachability in AR-to-MN direction.

D.1.2 Router Discovery

Router Discovery is achieved through the receipt of a router advertisement sent from the new AR. This will either be in the form of a router advertisement sent periodically to the all-nodes multicast address, or in response to a router solicitation sent by the MN. There is a potential race condition here; the MN will send a router solicitation *if* it discovers that its current AR is considered unreachable (i.e. its neighbour cache entry is not set to REACHABLE), and will thus, receive a solicited router advertisement from the new AR, or it will receive an unsolicited router advertisement from the new AR as part of its periodic broadcasts.

There is no guarantee as to which method will occur first. It depends on the exact circumstances at the time of handoff: (i) the period of router advertisement transmissions by the NAR and (ii) the exact value of the various timers at that moment in time. One may hypothesise that reducing the period of router advertisements will increase the likelihood of receiving an unsolicited router advertisement on the new link *before* realising that the PAR is no longer reachable. However, as we see from results of Section 3.7, the reduction of the router advertisement interval *does not* guarantee faster router discovery. Our results report a nominal hangover period of 84ms, while the router advertisement interval is configured at an average value of 40ms.

In addition, as noted earlier the receipt of a new unsolicited router advertisement is not necessarily a explicit indication of an IP network transition. Thus, the MN may also decide to confirm that its current AR is definitely unreachable *before* deciding to use the new AR. In the case of the IP handoff, this would translate into transmitting a number of neighbour solicitations for a pre-determined time without (i) receiving a corresponding neighbour advertisement from the current AR (transiting to previous state), (ii) utilising a periodic router advertisement from the new AR. This type of *numb* reactivity on the part of the MN, as prescribed by current protocol standards would yield a significant delay, when pursued during an IP handoff.

D.1.3 Care of Address Configuration

The MN must configure itself with an IPv6 address to be used on the new network. This will be the MN's New Care-of Address (NCoA). Address configuration can be performed in a stateful or a stateless manner. An IPv6 host may use both stateless and stateful address configuration completely independently from one another. The precise method to be used can be signalled with the setting of various flags in router advertisement messages.

If DHCPv6 (stateless or stateful) is to be used by the host for address configuration it incurs an extra overhead that is detrimental to expedient handoffs. DHCPv6 requires an extra request/response exchange on the new network in addition to normal router discovery mechanism.

Stateless Address Configuration

There are two ways in which an IPv6 node can configure its address in a stateless fashion:

- Using automatic address configuration with prefix discovery
- Using stateless DHCPv6

Automatic address configuration utilising prefix discovery is specified in [399]. If the 'autonomous' flag of a Prefix Information Option contained in a router advertisement is set, the IPv6 host may automatically generate its global IPv6 address by appending its 64-bit interface identifier to the prefix contained in the router advertisement. There are different ways in which the host may choose how to generate its interface identifier (e.g. based on MAC address, random or cryptographically generated). Description of such techniques is, however, beyond the scope of this investigation . Stateless DHCPv6 is not mentioned as an option given in router advertisements [399]. However recent discussions in the IETF IPNG WG have suggested signalling the usage of stateless DHCPv6 via the 'O' flag in router advertisements. At the time of writing the exact way of signalling that hosts should use stateless DHCPv6 is not clear. However, since there are few available implementations, this is not a major concern.

Stateful Address Configuration

With respect to IP handoff delay, using stateful DHCPv6 is not significantly different to using stateless DHCPv6 as the observed request/response time reported by work in [399] indicates the two being nearly similar in most cases.

D.1.4 Duplicate Address Detection

An MN must perform Duplicate Address Detection (DAD) when it bootstraps onto an IPv6 network to ensure that its configured care-of address (CoA) is unique on that link.

In IPv6, the DAD procedure is defined in IPv6 Stateless Address Auto-configuration [110], and uses the Neighbour Discovery procedures defined in [107]. An MN cannot use a new CoA until the DAD procedure has been successfully completed. During that period, the MN's new CoA is seen as *tentative*, and can be used solely for neighbour discovery purposes (of which the DAD procedure is part of). If an MN is allowed to use its new CoA *before* the DAD function is complete, while another node is using the same address on-link, the MN would erroneously process packets intended for the other host.

To perform DAD, the MN sends out a neighbour solicitation message with its own new CoA address as the target address of the solicitation message. The destination address in the IPv6 header of the neighbour solicitation is set to the solicited-node multicast address of the target address with the source address being the unspecified address. If there is another node on the link that is using the same address as the MN's new CoA, one of two things will happen:

- The node holding this IPv6 address will receive the MN's neighbour solicitation message and reply with a neighbour advertisement (sent to the all-nodes multicast address). In this manner, the MN is informed (and prevented from) configuring a duplicated IPv6 address.

- The MN will receive a neighbour solicitation with its new CoA as the target address from a competing host that is also in the process of performing DAD.

Thus, the DAD procedure will give an explicit indication to the MN should there be another node on the network that is using its new CoA. However, (and to the detriment of any node wishing to perform auto-configuration at haste) the DAD procedure provides no explicit indication that a MN's new CoA is not being used by another node on the network.

In fact, the point at which DAD can be considered to have succeeded is quite vague. According to [110], a node performing DAD can consider its tentative address unique if no indication of a duplicate address is observed within RETRANS_TIMER ms after sending DUP_ADDR DETECT_TRANSMITS number of neighbour solicitations.

Both the values of RETRANS_TIMER and DUP_ADDR DETECT_TRANSMITS are configurable parameters and by default are set to 1000 and 1 respectively. Therefore, under default conditions DAD is expected to take a minimum[2] of 1000 ms.

Note that [110] states that a node should *delay* sending its neighbour solicitation for DAD by a random time interval between 0 and MAX_RTR_SOLICITATION_DELAY seconds if it is the first packet sent from the interface after (re)initialisation

In [107], MAX_RTR_SOLICITATION_DELAY is defined as being 1000 ms in duration. Therefore, unless the MN has previously sent a router solicitation, it will incur further delay during its auto-configuration process. In the average case this will be an *extra* 500 ms, and up to an additional 1000 ms in the worst case.

D.1.5 Care-of Address Registration

Once the MN has detected that it has moved networks, obtained a new CoA and has been granted access to the network, it must inform its HA (Home Agent) of its new location. During the period from the moment the MN lost connectivity with its previous AR until the moment it informs its HA of its new location, all packets that have been sent to it will have been *lost*, while the MN is unable to send packets towards any of its CNs.

The MN registers its new CoA with its HA by sending a binding update (BU). The HA acknowledges this by replying with a binding acknowledgement (BAck) and is then able to tunnel packets bound to the MN's home address (HoA) towards MN's new location (i.e. MN's new CoA).

D.1.6 Binding Update Completion

This stage refers to the MN informing all CN peers of its new reachable location at its new CoA.

In a similar fashion to a HA binding update, the MN sends a BU to each CN. However, to protect against redirection attacks, a CN binding update is subject to an authentication procedure known as a Return Routability (RR) test. Such test as an authentication step whereby the CN receiving MN's binding update, can confirm its

[2] plus any additional delay for link transmissions and logic computation.

authenticity and thus exclude the possibility of malicious attack. An in-depth discussion of RR is out of scope for this study; a more detailed elaboration of RR signalling is found in [400].

In brief, RR uses a Home Test (HoT) and a Care-of Test (CoT). The tests are initiated by the MN but conducted by the CN (on the return path). The CN issues the two tests to the MN via the HA and the route-optimised path (i.e. direct to the new CoA) respectively. The MN replies with the answer to the two tests in the BU message sent to the CN. If the tests have been responded to (by the MN) correctly, the CN acknowledges the authenticity of MN's BU.

Once the MN has received the Binding Acknowledgement (BAck) from its CN, the handoff process is complete. Where route optimisation is not possible the IP handoff is complete once the new CoA has been registered with the HA.

D.2 Delay anatomy of the MIPv6 handoff process

D.2.1 Movement Detection time t_d

Movement detection time, denoted by t_d, is defined as the sum of two individual latency components:

- link-switching delay T_{l2}: this is the time delay pertaining to attachment of the MN with new PoA at the link-layer. With respect to 802.11 this refers to the (re-)association of the wireless station with the Access Point (AP) serving the new PoA.

- link-local IPv6 address configuration delay $T_{llconfig}$: this is the time between the first time that the MN encounters a new link by receiving neighbour adverts over its all nodes or solicited-nodes multicast address and configuration of a link-local address. Configuration of a link-local address is effected as well as L2 information is exchanged with the new AR.

The movement detection time can thus be expressed as:

$$t_d = T_{l2} + T_{llconfig} \tag{D.1}$$

D.2.2 IPv6 CoA Configuration Time t_c

We define CoA configuration time (t_c), as the time commencing from the moment of the receipt of a router advertisement (including the router advert solicitation if so used) to the moment that Duplicate Address Detection and the update of the routing table has completed. Depending on the mechanism employed to configure an IPv6 CoA, t_c may vary. However, the generic form of CoA configuration delay component is of the form:

$$t_c = T_{prefAdv} + T_{AddrConfig} + T_{RouteUpdate} \tag{D.2}$$

where $T_{prefAdv}$ is defined as:

$$T_{prefAdv} = \begin{cases} T_{rtAdv} - T_{rtSol} & \text{if rtAdv is solicited,} \\ \frac{rtAdvInterval}{2} & \text{(avg.) if rtAdv periodic} \end{cases} \tag{D.3}$$

For *stateless* IPv6 address auto-configuration [110], $T_{AddrConfig}$ denotes the time required *by the MN* to employ an address configuration rule to produce a unique, globally routable IPv6 address, as shown in Figure 3.5. For instance if the EUI64 address configuration rule is employed, the $T_{AddrConfig}$ delay component becomes:

$$T_{AddrConfig} = T_{EUI64} + T_{DAD} \tag{D.4}$$

We anticipate that for the stateless case, T_{EUI64} or other address configuration rule employed on the MN, that requires no protocol interactions with another host, should be dependent on the processor speed of the MN; as such the address configuration rule delay may be negligible compared to the total t_c.

T_{DAD} is the time required to resolve uniqueness of the configured IPv6 CoA. The mechanism to effect this is typically address resolution by transmitting a Neighbour Solicitation for this address to the all-nodes multicast address and then waiting for RetransTimer interval (Default 1000ms) before transmitting up to DupAddrDetectTransmits (Default 1). If during or after RetransTimer interval there has been no Neighbour Advertisement on the particular tentative CoA, the address is assumed to be unique

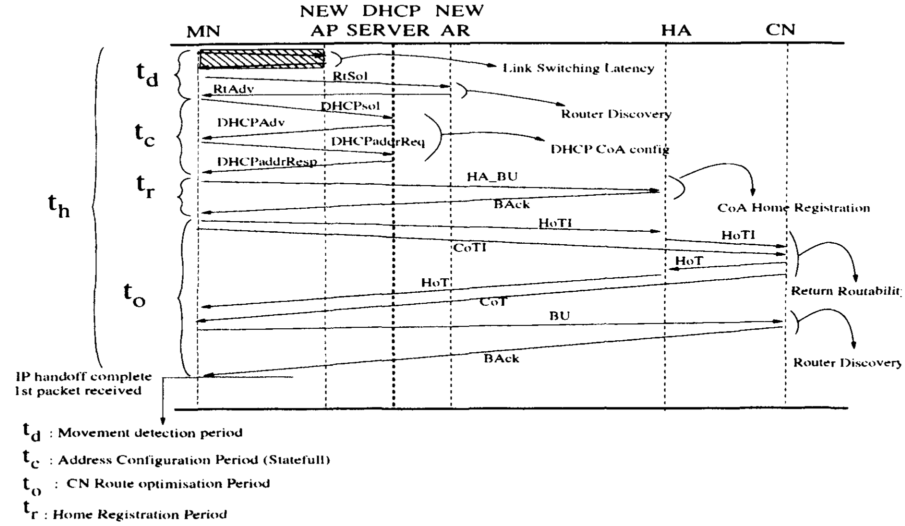


Figure D.1: IP handoff delay incurred under *Statefull* (DHCPv6) address auto-configuration by MIPv6 signalling

and assigned to the interface.

For the purposes of comparison we assume use of the default values in Mobile IPv6 since these are expected to be the standard default configuration values.

Thus, the T_{DAD} delay component is algebraically represented as:

$$T_{DAD} = T_{NeiSol} + \sum_{k=1}^{DupAddrDetectTransmits} T_{RetransTimer} \tag{D.5}$$

In the event of *stateful* address auto-configuration [401] the time for CoA configuration becomes:

$$T_{AddrConfig} = T_{DHCPsolicit} + T_{DHCPadvert} + T_{DHCPreq} + T_{DHCPreply} \tag{D.6}$$

$T_{DHCPsolicit}$ and $T_{DHCPadvert}$ denote the round trip time (RTT) required to solicit a DHCPv6 advertisement. $T_{DHCPaddrReq}$ and $T_{DHCPaddrResp}$ denote the second RTT delay incurred by requesting and acquiring an IPv6 address by the DHCPv6 server. Hence, the delay overhead incurred by these two signalling handshakes represents the total transmission delay incurred by stateful configuration of a CoA via a DHCPv6 server.

Figure D.1 illustrates the MIPv6 handover procedure where DHCPv6 is used for CoA configuration.

We may note that a DHCPv6 server is *not* necessarily available on the same link of the IPv6 handoff; hence, the delay component incurred by each DHCPv6 signal is tracked by the end-to-end delay between the DHCPv6 server and the MN.

Note it is quite likely that DHCPv6 is used even in the case of the MN using stateless address configuration as instructed by the received Router advertisement in Neighbour Discovery [107]; for example, a list of local DNS or NTP servers can be provided by the DHCPv6 server. We assume that such signalling exchanges do not affect the handoff latency as long as they are done 'out of band'[3] with respect to the address configuration task. The case where DHCPv6 is not used explicitly for address configuration purposes, but may affect in sequence of events the total handoff latency, is out of scope in this study.

DAD is more critical from a security perspective as opposed to accidental configuration of an IPv6 address already configured in that link (duplicate). While the latter is statistically very rare[4] with a probability of $1/2^{64}$, the former is a certain denial of service security risk for the visiting MN resulting into a false address resolution and reachability that is guaranteed to disrupt MN's active IP service sessions. Such disruption may manifest itself as through malicious address hijacking either as *connection hijacking* in the case of connection- less transport protocols or *connection resetting* in the case of connection-oriented (TCP) applications. It is thus, imperative that for stateless address configuration purposes the DAD process is enforced.

D.2.3 CoA Registration time t_r

The CoA registration time (t_r) is defined as the transmission delay incurred during registration of the MN CoA with its HA. This is essentially the RTT between the MN and HA plus associated processing of the BU and BA messages.

$$T_r = RTT_{MN-HA} + BU_{proc} + BAck_{proc} \tag{D.7}$$

D.2.4 Route Optimisation Time t_o

The route optimisation time (t_o) is defined as the transmission delay incurred during registration of the MN bindings with its corresponding peer. During communications with multiple CN's, such transmission delay is accounted by the delay incurred during

[3]asynchronously

[4]assuming a unique interface ID allocation algorithm like EUI64

signalling interactions with the CN with the longest RTT from the MN.

Depending on the mode of security effected in the BU registration process, there are two forms of binding updates: (i) unauthenticated and (ii) authenticated.

Unauthenticated BU

In the event of unauthenticated BU, the route optimisation time (t_o) is defined as the time period between a BU dispatched to the CN and the first data packet received by the MN from the CN. The BAck signal from the CN is typically piggybacked in the first data packet.

$$T_o = \begin{cases} RTT_{MN-CN} + BU_{proc} + BA_{proc} & \text{if BU not authenticated,} \\ T_{RR} + (RTT_{MN-CN} + BU_{proc} + BA_{proc}) & \text{if BU authenticated} \end{cases} \tag{D.8}$$

Authenticated BU

In the event of an authenticated BU using return routability (RR), the MN must first initiate the Home Test (HoT) and Care-of Test (CoT) before it can send a binding update to the CN. The RR procedure is illustrated in both Figure 3.5 and Figure D.1. The idea of these tests is that the MN 'proves' the authenticity of its network bindings (home and visited) by supplying proof that it received security data (keygen tokens) that the CN sent to these bindings. Such 'proof' ensures the authenticity of the upcoming BU signal dispatched to the CN.

The T_{RR} delay component is tracked by the RTT between the MN and CN and the sum of RTTs between the MN and HA as well as the HA and the CN. It has the following algebraic form:

$$T_{RR} = \begin{cases} RTT_{RR} + 2(T_{HoTIproc} + T_{HoTproc}) & \text{if } RTT_{RR} > RTT_{MN-CN}, \\ RTT_{MN-CN} + CoTI_{proc} + CoT_{proc} & \text{if } RTT_{RR} < RTT_{MN-CN} \end{cases} \tag{D.9}$$

where RTT_{RR} is the composite round trip time of a combined HoT-CoT signalling handshake expressed as:

$$RTT_{RR} = RTT_{HA-MN} + RTT_{HA-CN} \tag{D.10}$$

The above imply an additional RTT for the purposes of securing an authenticated BU subsequently sent to the CN.

Note that once the HA has acknowledged the registration of the MN's new CoA, any new CN attempting communications with the MN will succeed due to the HA being able to tunnel packets destined for the MN to its new CoA (i.e. using triangular, non-optimised routing). However, communication with any existing CN at the time the handoff occurred and with whom route optimisation was being used, cannot resume until the MN has successfully registered its new CoA with it by the RR procedure.

We emphasise that the frequency of RR establishment is dependent on the lifetime of tokens generated to protect the authenticity of the MN sending the BU; that is, under certain MIPv6 parameter configuration the RR process need not be applied on a per handoff basis. Token lifetime is controlled by timer parameters are MAX_RR_BINDING_LIFETIME and MAX_TOKEN_LIFETIME. These parameters, however, have 0 as their default value implying that the RR process by default is applied on a per handoff basis. For the purposes of this study we choose to analyse protocol behaviour under default parameter settings.

D.3 Calculating jitter in experimental measurements

Jitter is calculated through an autoregressive moving average, according to the specification of ??. For the purposes of our measurements investigation, jitter is calculated as follows:

$$J(i) = J(i-1) + \frac{|D(i-1,i)| - J(i-1)}{16} \tag{D.11}$$

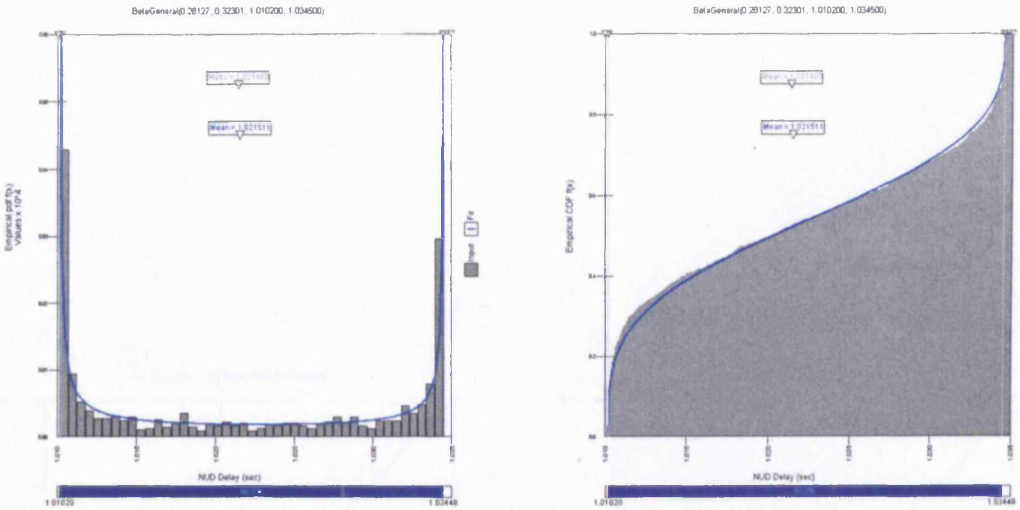
where difference $D(i-1,i)$ in packet spacing between packets p_{i-1} and p_i may be derived from :

$$D(i,j) = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i) \tag{D.12}$$

where S_i is the (sender[5]) RTP time-stamp from packet i, and R_i is the (receiver) time of arrival in RTP time-stamp units for packet i. Jitter is calculated off-line from traces obtained from packet captures obtained.

[5]It is intuitive that $(S_j - S_i)$ identify the Packetisation rate at the sender (e.g. r_{GSM} = 20ms)

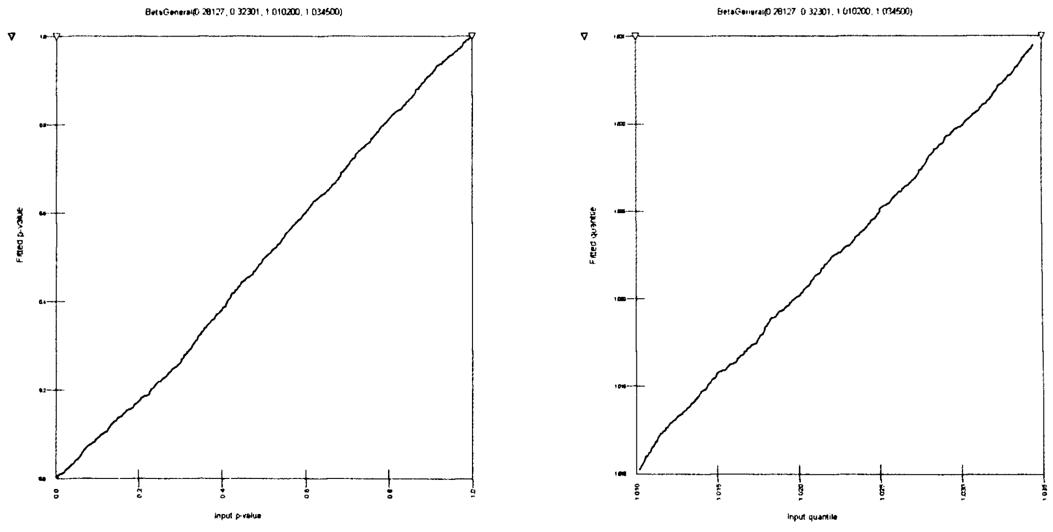
D.4 Neighbourh Reachability delay distributions during an h2v MIPv6 handoff



(a) Probability Density Function

(b) Cumulative Distribution Function

Figure D.2: MIPv6 handoff delay component induced by neighbour reachability signalling on-link

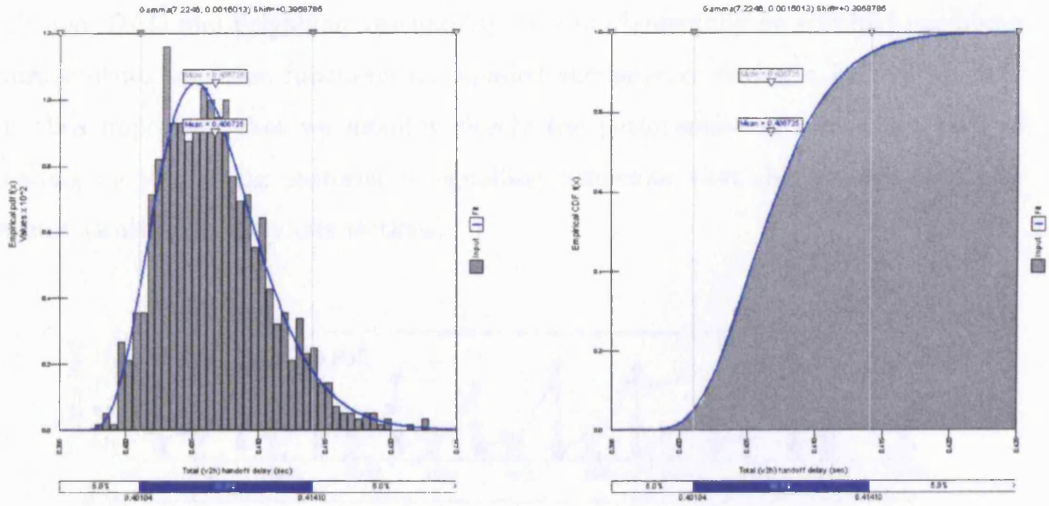


(a) Probability plot

(b) Q-Q plot

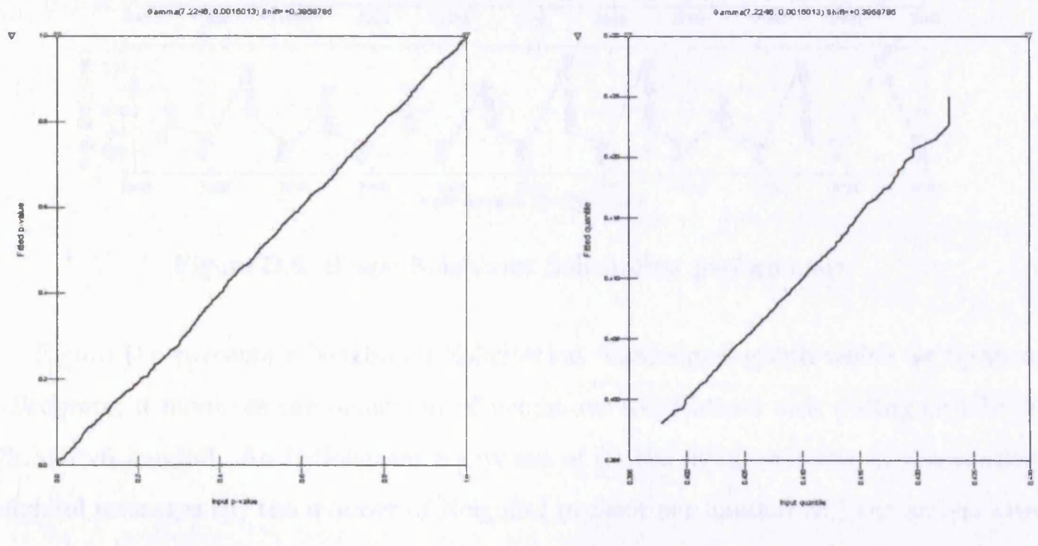
Figure D.3: P-P and Q-Q plots for latency induced by neighbour reachability signalling during a MIPv6 handoff

D.5 Statistical distribution of delay during an v2h MIPv6 handoff



(a) Probability Density Function (b) Cumulative Distribution Function

Figure D.4: Distribution of total (v2h) MIPv6 handoff delay for the MN returning to the home network



(a) Probability plot (b) Q-Q plot

Figure D.5: P-P and Q-Q plots for the derived distribution applicable to total (v2h) MIPv6 handoff delay

D.5.1 MIPv6-specific Neighbour Discovery Performance

Neighbour Solicitations/Advertisements

We have seen that the three key functions of IPv6 Neighbour Discovery, namely, address resolution, DAD and neighbour reachability, rely fundamentally on solicited neighbour advertisements, as these functions are applied successively during a MIPv6 handoff. It is thus important that we monitor closely the performance of such class of IPv6 signalling by identifying patterns of signalling behaviour that characterise the delay measure identified in previous sections.

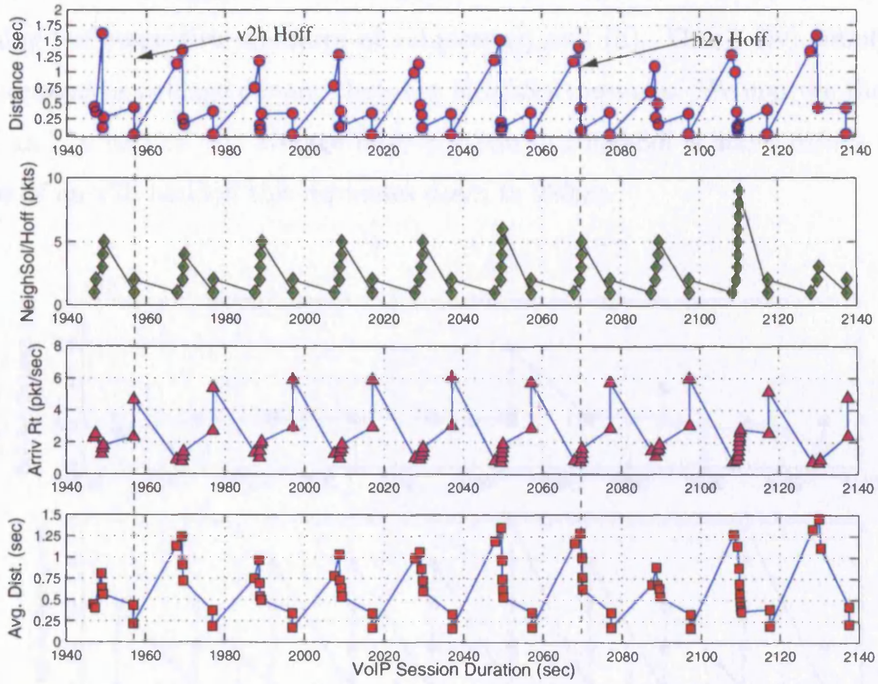


Figure D.6: Single Neighbour Solicitation performance

Figure D.6 presents a Neighbour Solicitation 'vital signs' graph which we term as *N-Soligram*; it monitors the behaviour of neighbour solicitations sent during an h2v or v2h MIPv6 handoff. An N-Soligram comprises of (i) the distance between consecutive NeighSol messages (ii) the number of NeighSol packets per handoff (iii) the arrival rate NeighSol messages (iv) the instantaneous average distance between router solicitations. In all four subgraphs the monitoring is confined to the critical period of the handoff which is the focus of our investigation. Furthermore handoffs alternate between visited (h2v) and home (v2h) networks.

In the case of (i) an h2v MIPv6 handoff is identified by a high distance 'peak', whereas v2h handoff is described by low distance spikes. The high distance peak is justified by the delay incurred by the DAD process and the subsequent delay incurred by NUD. Subplot (ii) show the frequency of NeighSol messages; clearly an h2v handoff incurs more NeighSol messages (around 5) than a v2h handoff (only 1). This is because a visited network (AR) has no information about the address[6] of the MN or its reachability, in contrast to the home network whereby the HA constantly 'defends' the MN's home address.

Subplot (iii) shows a higher arrival rate of NeighSol message in the v2h handoff case. This is justified by the fact that a v2h handoff completes significantly faster and thus the two NeighSols are sent within a smaller period of time; this is clear by combining the respective measures of subplots (i) and (ii). Graph (iv) simply shows the instantaneous average distance between NeighSol messages. We may see that in the case of an h2v handoff the average interval between NeighSol is about 659ms while in the case of an v2h handoff this decreases down to 280ms.

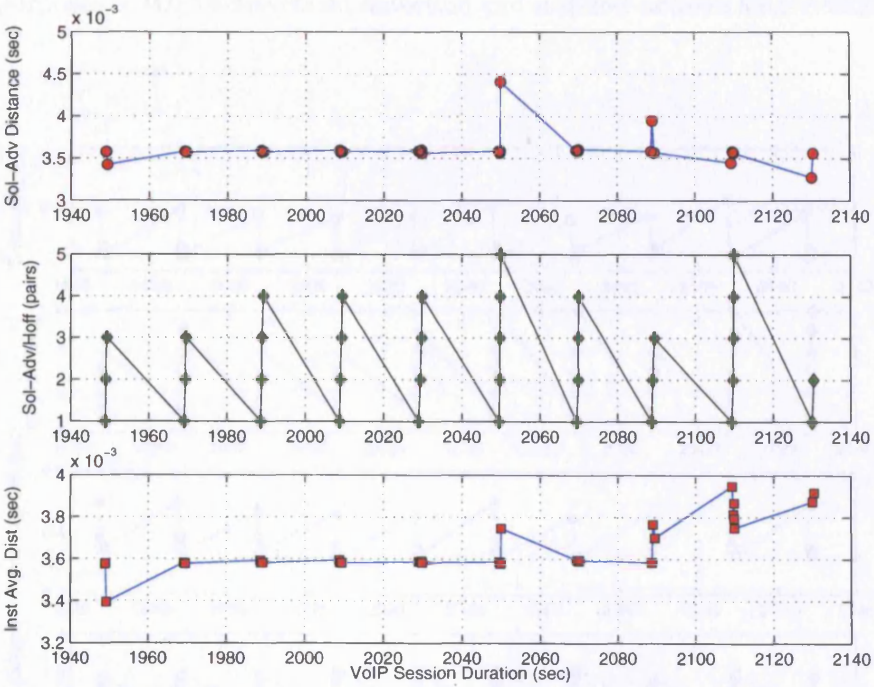


Figure D.7: Solicited Neighbour Advertisement performance

The second N-Soligram of figure D.7 shows the performance of a solicited NeighAdv message pair, which we term as *N-SolAgram.* An N-SolAgram comprises of (i) the distance between NeighSol and NeighAdv messages, (ii) the number of solicited advert

[6]MAC address and uniqueness of the IPv6 CoA address

message pairs, (iii) the instantaneous distance average between the two messages. Both (i) and (iii) agree that the distance of a NeighSol from a NeighAdv message is about 3.6ms, while (ii) confirms that the actual number of solicited NeighAdv messages is 3-4. The arrival rate of these message pairs is tracked by the arrival rate of NeighSol messages shown in figure D.6.

From the above it becomes obvious that while the solicited NeighAdv message incur insignificant delay to the MIPv6 handoff process, *the distance of successive NeighSol messages result into significant delays during the MIPv6 handoff*, given that different NeighSol messages signify different functions of the Neighbour discovery process. This is particularly important during the functions of DAD and NUD which account for the majority of NeighSol/NeighAdv signals during a MIPv6 handoff.

Router Solicitations/Advertisements

In a similar fashion to N-Soligram and N-SolAgram plots we produce the respective R-Soligram and R-SolAgram monitor plots. Solicited router advertisements are essential for the purposes of MIPv6 movement detection and stateless address auto-configuration.

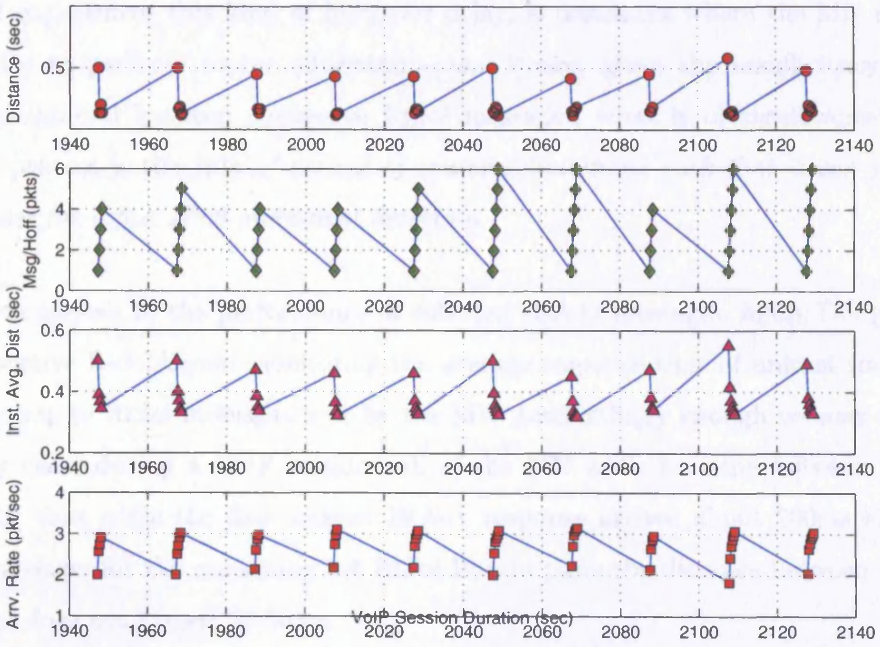


Figure D.8: Single Router Solicitation performance

The R-Soligram of figure D.8 shows the performance of single router solicitations for (i) inter-RtSol distance during a handoff, (ii) no of RtSol messages per handoff, (iii)

RtSol arrival rate during a handoff (iv) instantaneous average distance. The monitor graph focuses solely on h2v handoff since a v2h MIPv6 handoff requires no router solicitation; it simply employs the RtAdv message arriving periodically to the MN from the home network AR.

We may observe that the mean distance between RtSol messages is 378ms. This is offset by the first router solicitation which is separated by the second one by a time distance of 500ms with subsequent RtSols sent every 250ms. While this is faster than specification (RtSol interval is 1000ms), to make provisions for faster movement detection, it proves to be insufficient, given the amount of RtSol messages sent by the MN in search of a RtAdvert before address configuration and subsequently DAD process gets initiated.

Furthermore, we have observed that the function of soliciting a router advertisement, characterising collectively the movement detection process, consumes 1-3 RtSol messages before a CoA is configured and DAD function is initiated. The precise number of RtSol messages depends on whether an RtSol message has been lost during the type of hangover delay experience by the MN during the MIPv6 handoff.

Given the distance observed between RtSol messages it comes as no surprise why the MN experiences this kind of hangover delay, at instances where the MN remains insensitive to periodic router advertisements. Hence, given the insufficiency of the observed interval between successive RtSol messages, what is of significance to the handoff process is the *rate of arrival of router solicitations such that it can provided guarantees for a fast IPv6 movement detection.*

With respect to the performance of solicited RtAdv messages, figure D.9 presents the respective R-SolAgram monitoring the average response time of unicast router advertisements to RtSol messages sent by the MN. Interestingly enough we may observe in many cases during a VoIP session call of the MN while roaming between WLAN networks, that while the first unicast RtAdv response arrives about 200ms after the RtSol message, for the remaining 3-4 RtSol-RtAdv pairs the distance between the two messages does not exceed 50-60ms.

The latter while significantly lower than the specified random delay interval of 0-500ms imposed by the Neighbour Discovery specification, it induces a non-negligible delay component that fast detracts from the golden 200ms VoIP one-way delay budget pertaining to interactivity in the communication pattern between the two peers.

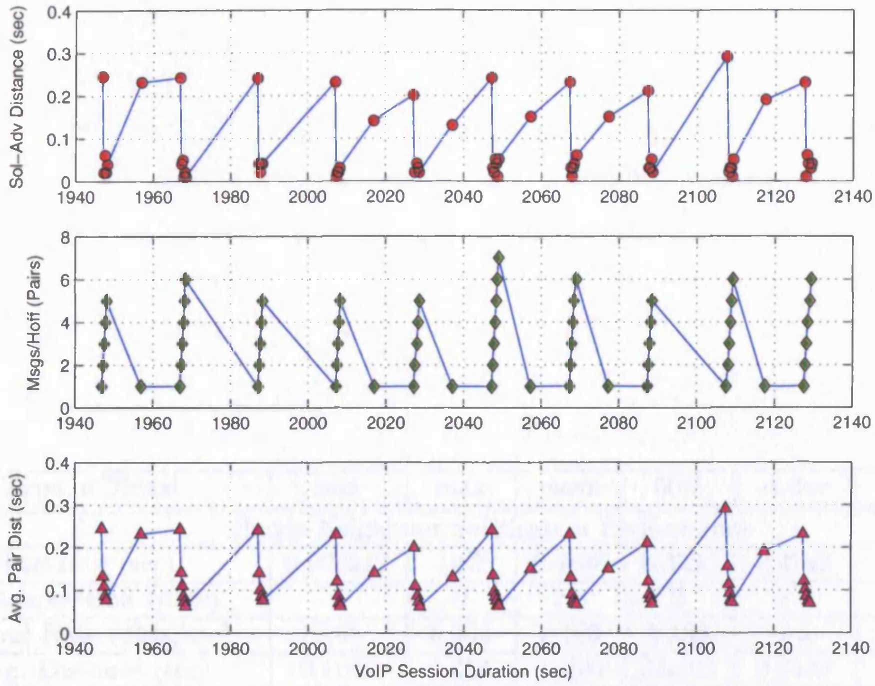


Figure D.9: Solicited Neighbour Advertisement performance

With respect to the first solicited RtAdv pair, it should be noted that the delay peak is experienced due to 802.11b link-switching. That is to say, such delay spike is not caused by the delay on the part of the router to respond, but because during the L2-handoff MAC retransmission attempt to deliver the frame to the new AP and subsequently to the new point of AR attachment.

Also it appears from the above, that the boundaries of 0-500ms random delay imposed by the ND specification for solicited routers advertisements is a rather rare event and thus does not affect the handoff process as reported in [402, 403], according to the results derived by our experimental measurements dataset over a protocol-compliant MIPv6 implementation.

Table D.1 provides a summary of the first statistical moments on Neighbour discovery performance with respect to Neighbour solicitations, solicited Neighbour adverts as well as router solicitation and solicited router advertisements. The performance of these signals summarise the rate of availability of control signalling with respect to neighbour discovery for MIPv6 mobility management purposes. The statistics collected focus on the h2v case of a MIPv6 handoff since the v2h case consumes very little ND signalling.

Type of Signal	min	max	mean	50%	st.dev	90%	99%
Single Neighbour Solicitation Performance							
Distance (sec)	0.000212	1.62	0.4596	0.3298	0.4742	1.273	1.611
Messages/Hoff (pkts)	1	9	2.8	2	1.8	5	8.78
Arrival Rate (pkts/sec)	0.68	6.065	2.166	1.589	1.5	5.213	6.04
Avg. Distance (sec)	0.1649	1.455	0.6591	0.6292	0.3456	1.176	1.431
Solicited Neighbour Advert Performance							
Sol-Adv Distance (sec)	0.0032	0.044	0.0035	0.0035	0.00016	0.0036	0.0044
Msg Pairs/Hoff (pair pkts)	1	5	2.4	2	1.2	4	5
Avg. Distance (sec)	0.0033	0.0039	0.0035	0.0036	0.00012	0.0039	0.0039
Single Router Solicitation Performance							
Distance (sec)	0.27	0.55	0.3294	0.295	0.0806	0.48	0.55
Messages/Hoff (pkts)	1	6	3.08	3	1.563	5	6
Arrival Rate (pkts/sec)	1.81	3.14	2.698	2.83	0.3629	3.07	3.14
Avg. Distance (sec)	0.3183	0.55	0.378	0.3533	0.0593	0.48	0.55
Solicited Router Advert Performance							
Sol-Adv Distance (sec)	0.01099	0.291	0.0778	0.041	0.0827	0.2313	0.2854
Msg Pairs/Hoff (pair pkts)	1	7	3.113	3	1.738	5.3	6.88
Avg. Distance (sec)	0.0627	0.291	0.1228	0.0993	0.0604	0.2138	0.2731

Table D.1: Statistical moments of Neighbour discovery signalling during h2v MIPv6 handoff

D.6 Influence of the Wireless medium

So far we have seen that the L2-handoff as effected at the MAC layer of 802.11b MAC/PHY protocol incurs a significant (above 200ms) latency component, which is by itself capable of impeding the interactivity of VoIP communications.

While we proceed to analyse the behaviour or an L2-handoff over an 802.11b (WLAN) interface, it is worth noting that observations from this section offer little space to generalisation for all wireless interfaces. Reason for that is the fact that wireless technologies differentiate from one another due to the disparity of medium access control (MAC) mechanisms supported by each one. Such disparity is justified by different design considerations and decisions made for each wireless protocol specification.

For instance the wireless medium (MAC+PHY) of WLANs differs fundamentally from cellular protocols such as GSM/GRPS or UMTS with most of the observations made being non-applicable for cellular protocols. This is because the initial specification of WLANs is geared towards short-range packet-switched communications multiplexed over a single wireless carrier; on the contrary GSM or UMTS wireless specifications are geared predominantly towards long-range, low-bandwidth (compared to WLAN) circuit- or on occasions packet-switched communications over allocated frequency channels.

Furthermore, a WLAN does not require per-flow channel setup/establishment as all traffic is multiplexed over a single carrier onto a deregulated frequency domain. On the contrary, GPRS and UMTS technologies require frequency planning/allocation on a per-subscriber basis, as each subscriber negotiates the allocation of a channel during call setup time. In addition, IPv6 control signalling is not effected over native IPv6 signals but interfaces over the GPRS/UMTS signalling stack. This is one of the main reasons why IPv6 handoffs over cellular networks require significantly more time to complete over native GPRS/UMTS signalling, as shown in [404, 293].

Nevertheless, while our wireless medium observations may be of limited value for every wireless technology, they provide in-depth guidelines on trends of wireless performance in the constantly growing family of IEEE802.1x WLANs, characterised by deregulated frequency domains where no frequency band licensing is required. In this light, (circuit and/or packet-switched) cellular (or hybrid) wireless interfaces such as GPRS or UMTS are out of scope in this part of investigation.

D.6.1 The IEEE802.11 link-layer handoff process

The handoff process employed by the link layer (L2) of the MN during its roaming between two neighbouring 802.11 WLAN is identified by 4 distinct processes as shown in

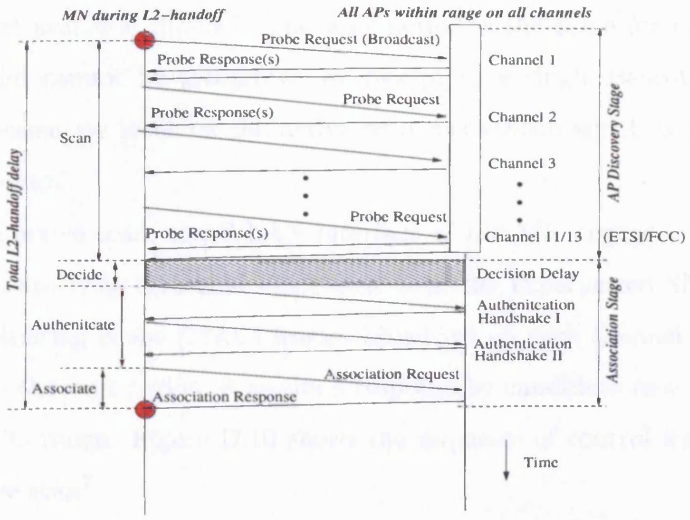


Figure D.10: The WLAN (IEEE802.11) L2-handoff process

figure D.10 : (i) de-association (ii) AP discovery (iii) Authentication (iv) re-association.

Depending on the signalling rate that the WLAN MN may be operating, the signal strength and signal to noise ratio (SNR) of the waveform (modulated digital frame) communicated with the AP, is expected to degrade due to path loss effects or attenuation factors [29] as the MN roams away from the AP. Such signal degradation may be manifested by obstructions in the line of sight between MN and AP [405] or attenuation factors such as trees, geo-climatic conditions etc.

In such case, the sensitivity threshold associated with the SNR experienced, is exceeded and the bit error rate (BER) increases beyond acceptable levels for communication purposes, to the extent that neither MAC retransmissions nor error correction techniques [406, 407] are capable of recovering lost information in the received MAC frame at the MN or AP.

At this point the MN (WLAN station) needs to discover other APs within its range (bound by the lowest signalling rate), amongst which it may choose one to associate. To this effect the MAC sublayer of the 802.11 protocol at the MN, engages into a *AP discovery* phase. The 802.11 specification defines two types of AP discovery: *passive* or *active*.

During a passive scan, the WLAN interface of the MN listens for *beacon* MAC protocol data units (MPDUs), that announce the availability of an AP. Beacons are sent out periodically by the APs at the default rate of 100ms on their pre-assigned channel. The MN awaits at each individual channel a fixed period before passively sensing the next available channel. The wait period is the same for both passive and active scans and cannot be preempted by receipt of a single Beacon frame for one AP. For this reason we focus on the active scan mechanism which is also part of our experimental setup.

During an active scan, the WLAN interface of the MN engages reactively (upon exceeding the sensitivity threshold associated with the experienced SNR) to discover new APs by initiating probe (MAC) frames broadcast on each channel scanned in succession. During the wait period, it awaits a response by candidate new APs potentially available with its range. Figure D.10 shows the sequence of control frames exchanged during an active scan[7]

Upon completion of the discovery phase, the MN creates an ordered set of candidate APs ordered by received signal strength. *For a passive scan, this implies that the sole criterion of AP selection at the MAC layer is signal strength, with cascading effects onto both the IP and (VoIP) application layers.* That is to say, signal strength effectively decides which network, should the MN continue its 'roamed' VoIP call.

Reactive discovery of candidate APs allows the MN to actively seek candidate (or a particular) APs, while the selection decision may also be based on criteria other than signal strength, such as 802.11 network ID (ESSID). That is to say, *reactive AP discovery allows greater flexibility in selection criteria to the MN when associating with an AP.* This can affect positively the QoS or billing experienced by an MN as it roams among different visited WLAN last-hop networks. As we see, however, in following sections and witnessed already in previous ones, the reactive character of AP discovery imposes significant delay so as to impede interactivity and real-time guarantees to multimedia services such as VoIP calls.

Figures D.11 and D.12 illustrate the breakdown of the L2-handoff component from the total MIPv6 handoff delay described in previous sections. It can be seen from figure D.11 that a set of 5-6 Probe Request messages are being broadcast before the first one that is responded to by the new AP of the visited network. This set of messages, actually, receive a response by the AP of the home network (and thus the response in this set of probe requests is not picked up by the sniffer monitoring the visited WLAN network).

On the contrary, as the MN sweeps the remaining 5-6 channels with probe requests,

[7]in a passive scan the probe request cease to exist while the probe requests are replaced by periodic beacons

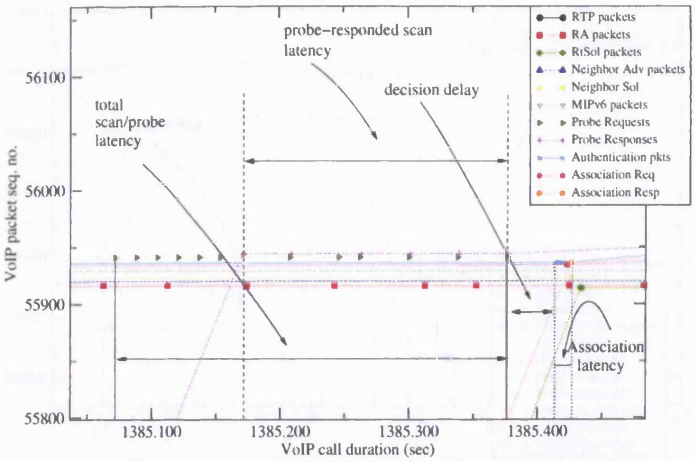


Figure D.11: Total (h2v) L2 handoff delay

we can see that the new AP neighbour at the visited network provides a probe response to all requests despite the fact that each request is sent to a different 'channel'. This is owed to a *channel leaking* effect that is characteristic of the 802.11 specification at the physical (RF) layer. We describe the effect and its importance to MIPv6 handoffs in the following section.

It is noted, that the devised 802.11 link-layer measurement setup exploits such channel leaking effect [408] in the frame analyser employed in the ensemble of the two WLAN monitor hosts[8] , for the purposes of *eliminating* capture losses in the frame traces recorded during each the VoIP call. Our measurements eliminate the loss percentage imposed by the sniffing process, in contrast to similar measurements conducted in [409].

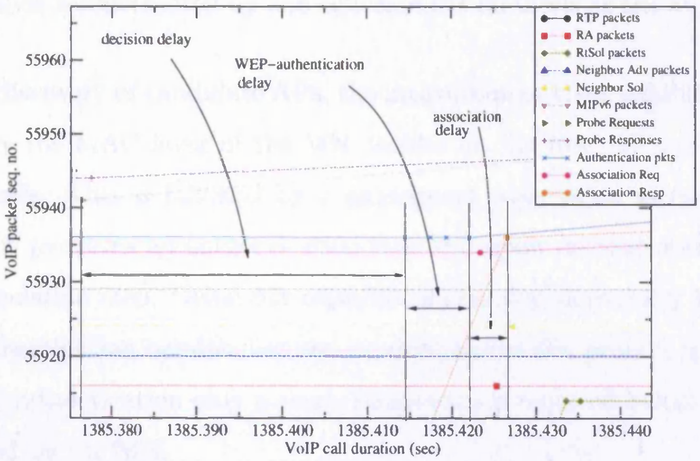


Figure D.12: (h2v) Association latency component

[8]We have used AiroPeek NX kindly provided by WildPackets Inc. for monitoring robustly the 802.11 interface at all signalling rates

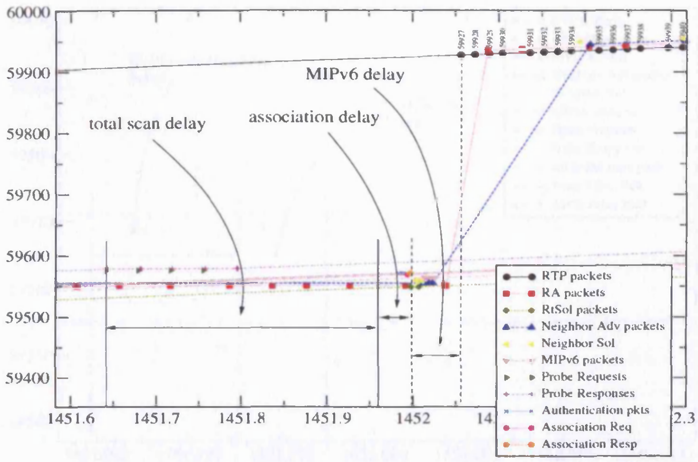


Figure D.13: Total L2-handoff delay for the MN returning back to the home network (v2h MIPv6 handoff)

It is found that the average number of Probe request messages varies between 8 and 13. While this is not characteristic of all 802.11 vendor implementations [303] it signifies that out of the 13 available channels (ETSI spec.) almost all are being searched sequentially before an association decision is made, with no preemption by the first available AP candidate. *This has a negative impact on both L2-handoff delay and subsequently the total MIPv6 handoff delay and packet loss imposed during a VoIP call.*

The impact of delay incurred by such design decision at the AP discovery phase is more pronounced at the v2h MIPv6 handoff than the h2v handoff, since in the latter case the MIPv6 delay contribution is far greater than the L2 handoff delay. Nevertheless, both delay components (in either case of MIPv6 handoffs) incur significant packet loss (> 90) that cannot reconstructed by loss concealment methods at the application layer.

Following discovery of candidate APs, the measurement trace exhibits a small delay period, whereby the MAC layer of the MN decides on the best AP available in terms of signal strength. This is followed by a subsequent association latency component, whereby the MN performs an authentication step and upon successful authentication a subsequent association step. Given our experiments employ shared-key WEP authentication, two authentication handshakes are required before the process is complete. For open key (null) authentication only a single handshake is required before an association request is placed by the MN.

Figures D.13 and D.14 show the breakdown of the v2h L2-handoff component contrasted by the MIPv6 delay incurred by signalling exchanges by both Neighbour

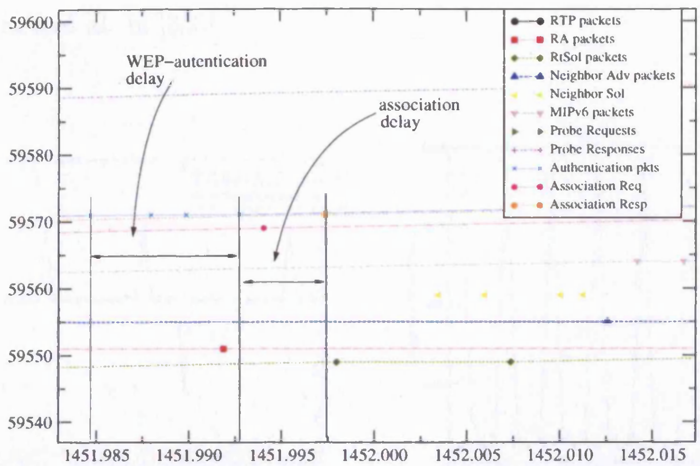
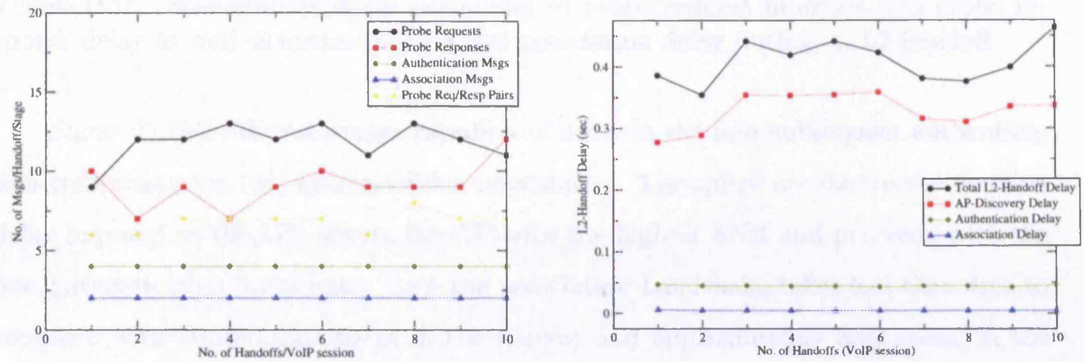


Figure D.14: Association latency component for the MN returning back to the home network (v2h MIPv6 handoff)



(a) No. of MAC signals

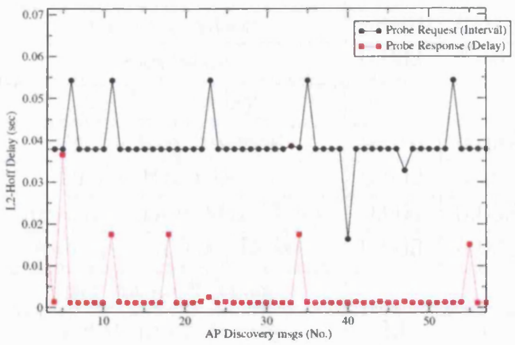
(b) L2-Handoff Delay components

Figure D.15: Number of messages during the L2-handoff process and delay components comprising the total L2-handoff latency

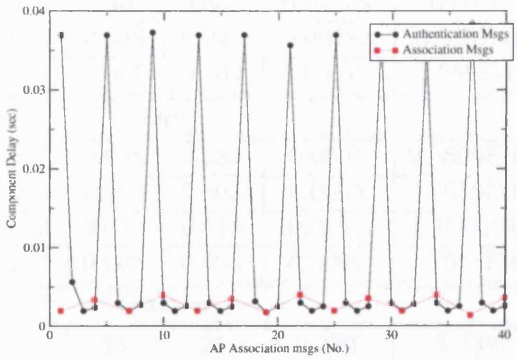
Discovery and MIPv6 protocols. It can be seen that *the L2-handoff delay is significantly greater than the actual delay incurred by the MIPv6 process when the MN handoffs back to the home network.*

Figures D.15(a) and D.15(b) show the breakdown of the discovery authentication and association delay components during an 802.11 handoff as part of the MIPv6 handoff process. The AP-discovery phase exhibits a delay that is logistically distributed ($\alpha = 0.3325$ and $\beta = 0.024864$) with a mean of 332ms and Std. Dev.= 0.04509.

Figure D.16(a) shows the delay incurred periodically while the probe request is sent out by the MN scanning the available channels for candidate APs. It can be seen that probe response take minimal time (around 3ms) in comparison to the probe request interval (around 38ms). These results are in agreement with independent results reported by Misra et al. in [303].



(a) AP Probe Request/Response delay



(b) Association-related delay

Figure D.16: Scan-specific delay pertaining to probe request intervals and probe response delay as well as authentication and association delay during an L2-handoff

Figure D.16(b) demonstrates variation of delay in the four subsequent authentication comprising the two authentication handshakes. The spikes are due to the decision delay imposed as the MN selects the AP with the highest SNR and proceeds with the first authentication handshake. Last the association handshake takes less than 4ms to complete with around 1ms spent in the request and approximately 3ms spend in the response by the AP.

Table D.2 provides a summary of the statistical moments characterising the individual components of L2-handoff delay at a data rate of 1 Mbps, as well as key delay measures such as the periodicity of probe request messages. These handoffs are effected during, both h2v and v2h MIPv6 handoffs which have been analysed individually in previous sections.

Table D.3 provides a summary of the statistical moments characterising the individual components of L2-handoff delay while both MN and AP communicate at a signalling rate of 11 Mbps, during both h2v and v2h MIPv6 handoffs.

It is interesting to observe that, *as the signalling rate increases the mean L2-handoff delay remains approximately constant, but the number of control frames increases*; this is particularly the case for the AP-Discovery phase which involves a significant number of L2 frame exchanges between the MN and an AP. This is counter-intuitive since as we see in Section F.8 a faster signalling rate driven by an underlying faster modulation scheme, should complete transmission of the same number of control frames in an L2

L2 Handoff	min	max	50%	95%	Mean	Std Dev	Variance
per handoff Delay	(sec)						
Total	0.248	0.536	0.407	0.501	0.423	0.08623	0.00246
Discovery	0.221	0.471	0.336	0.405	0.332	0.04509	0.00203
Authentication	0.040	0.046	0.044	0.046	0.043	0.00182	3.319E-06
Association	0.002	0.005	0.004	0.005	0.004	0.00079	6.359E-07
per Msg Delay	(sec)						
Probe Req Interval	0.016	0.092	0.037	0.058	0.039	0.0095	9.0809E-05
Probe Resp Delay	0.002	0.004	0.003	0.003	0.003	0.0005	2.9856E-07
Authentication Msg Delay	0.004	0.038	0.003	0.050	0.010	0.0150	0.0002238
Association Msg Delay	0.003	0.005	0.004	0.005	0.004	0.0004	1.7978E-07
per handoff Msgs	(No. of)						
Total no of frames	23	32	24	28	24	2.332	5.4397
Probe Requests	9	13	11	12	11	0.974	0.949
Probe Responses	6	15	7	9	7	2.3116	5.3436
Probe Pairs	5	8	5	7	6	0.6759	0.4568
non-paired Prob Req	3	6	4	3	4	0.809	0.654
non-paired Prob Resp	1	8	2	4	2	2.907	8.445

Table D.2: L2-handoff delay statistics while both MN and AP operate at 1 Mbps (P = 5mW, AP-AP and AP-MN distance = 3m)

handoff within a smaller amount of time and hence effect a faster L2-handoff.

Looking at the number of messages we observe an increase in the total number of messages exchanged at the MAC layer during the L2-handoff as captured by the 802.11 frame analyser.

A more careful examination of these frames reveals that a percentage of these frames are actually retransmitted[9] during the course of the L2 handoff. Looking back at the percentage of retransmissions at 1 Mbps signalling rate we observe that the number of retransmitted frames is significantly smaller during the respective L2-handoff. This attests that during a lower signalling rate the amount of ARQ retransmissions is significantly lower, since a lower signalling rate employs by design a more robust modulation scheme that packs fewer bits per symbol (e.g. BPSK at 1 Mbps) [51]. On the contrary for a higher signalling rate the amount of retransmission increases significantly as a result of a less robust modulation scheme packing more bits per symbol (e.g. CCK at 11Mbps).

Thus for a transmission range which is attainable for both signalling rate boundaries in 802.11b (i.e. 1 and 11Mbps) and for the same amount of SNR we may conclude that, under the same propagation conditions, *a lower signalling rate (1 Mbps) will experience*

[9] as attested by the relevant flags of the 802.11 header

L2 Handoff	min	max	50%	95%	Mean	Std Dev	Variance
per handoff Delay	(sec)						
Total	0.271	0.478	0.431	0.469	0.43	0.025	0.00064
Discovery	0.272	0.442	0.353	0.433	0.355	0.044	0.00196
Authentication	0.040	0.044	0.041	0.042	0.041	0.00044	1.9360E-07
Association	0.002	0.004	0.002	0.003	0.002	0.00031	1.0198E-07
per Msg Delay	(sec)						
Probe Req Interval	0.016	0.113	0.040	0.068	0.042	0.014	0.0002013
Probe Resp Delay	0.0004	0.0007	0.0004	0.001	0.002	0.02635	0.0006945
Authentication Msg Delay	0.0004	0.005	0.0017	0.020	0.010	0.01528	0.0002316
Association Msg Delay	0.0023	0.0043	0.0026	0.0037	0.0028	0.00047	2.2773E-07
per L2-Handoff Msgs	(No. of)						
Total control frames	30	45	36	43	38	3.5621	12.6885
Probe Requests	15	23	18	18	17	3.850	3.178
Probe Responses	8	16	10	15	14	2.9037	8.4317
Probe Pairs	5	8	6	8	6	0.714	0.5107
non-paired Prob Req	6	15	12	10	11	1.5653	2.4503
non-paired Prob Resp	4	8	4	7	8	2.5811	6.6621

Table D.3: L2-handoff delay statistics while both MN and AP operate at 11 Mbps (P = 5mW, AP-AP and AP-MN distance = 3m)

the same L2-handoff delay as seen at the highest signalling rate (11 Mbps). This is because the (propagation) speed of a faster modulation scheme at the highest signalling rate is diluted by an increased number of ARQ re-transmissions.

It can thus, be said that the lowest signalling rate effects indirectly, a proportionally faster L2-handoff than the highest signalling rate, given that the low data rate employs a slower modulation scheme than the faster data rate.

To contrast the propagation speed between high and low signalling rates we may compare signalling effected with one or two control frames at 1 and 11 Mbps. For instance, both probe response and authentication frames experience a transmission delay of approximately 400 usec at 11 Mbps; the respective signals experience a transmission delay between 2-4ms. This almost half to one order or magnitude more at 1 Mbps than at 11 Mbps.

AppendixD shows details of the distributions for the respective statistics shown in tables D.2 and D.3.

MAC/PHY layer intrinsics

In the previous section we noticed briefly the *channel leakage* effect that is characteristic of the 802.11 specification, at the physical layer. We describe briefly the reasons and design decisions of the 802.11 specification behind such effect while we also dis-

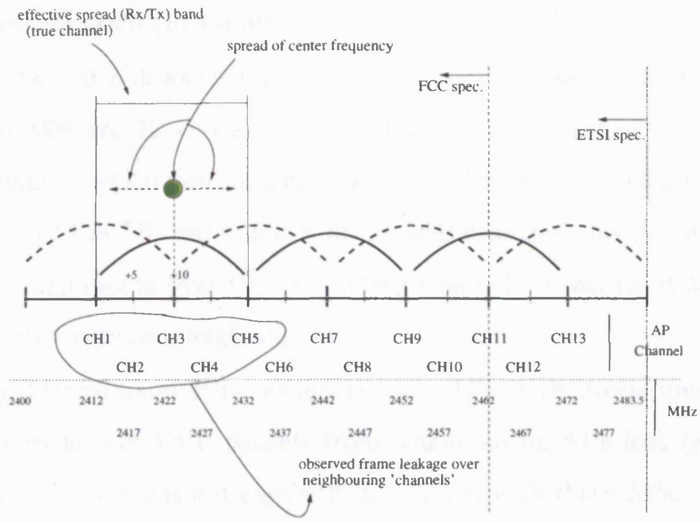


Figure D.17: Spread Spectrum Channel Overlap - European (ETSI) specification

cuss its implications over a dense multi-domain WLAN cloud effecting a geographically continuous coverage area for the MIPv6-enabled MN on the move.

The notion of 'channel' of IEEE802.11 WLANs, bears slightly misleading semantics when compared to the one of cellular communications. This is because in cellular communications a channel allocated to a mobile device maps uniquely to a discrete frequency band that does not overlap with other adjacent channels (enjoying a minimum frequency separation).

In 802.11, the channel over which the MN is associating (with an AP) and transmitting its VoIP frames, does *not* refer to a discrete, single frequency band. This is due to the nature of (direct sequence) spread spectrum (DSSS) in 802.11 signal modulation. The spreading implies that the actual RF signal energy is not constrained within a single discrete frequency; instead it is spread over a small frequency range. This is shown in figure D.17.

The above implies further that signal discrimination in 802.11, allows RF signals to slightly interfere with each other if an acceptable inter-channel separation is not maintained. In particular, the centre frequencies of each 'channel' are separated by 5 MHz but the signals are spread +/- 10 MHz from the centre frequency. Such design approach in 802.11 specification results into an intentional overlap with neighbouring discrete frequencies.

The number of configurable channels is dependent on regional specification of the 802.11 protocol; the European specification (ETSI) allows for 13 configurable channels, while the US specification (FCC) allows for only 11 channels.

From the above it follows that while the number of configurable channels on either the AP or the MN are 13 (ETSI) or 11 (FCC), the number of true channels with acceptable frequency separation, reduce down to 3; that is, in a 3-channel configuration the air interface of the MN experiences no interference from neighbouring APs, since the energy of transmissions over their operating channels is carried at a frequency that is clearly separated between neighbours).

By testing MIPv6-handoff performance over APs with overlapping channels, we find that transmissions of VoIP packets from neighbouring APs *leak* (get received by) to the MN, while the latter is not explicitly associated with these APs. This is because, during an MIPv6 handoff the VoIP packets gets delivered first to the MAC address of the wireless MN.

If the MN operates on the new AP at a channel overlapping with the AP of the previous network, then the MN can receive such packets, given that the MN has still configured its home address on its WLAN interface. Such effect may report significantly reduced[10] L3-handoff times [409, 105] that do not reflect accurately the true MIPv6 handoff performance shown in previous sections.

The only means to avoid such *channel leakage* onto the MN is to enforce an inter-AP channel separation[11] of $\geq 25MHz$.

From the above we may conclude *that performance of a MIPv6 handoff mechanism cannot be objectively measured unless sufficient frequency separation between the operating channels of neighbouring APs is guaranteed.* This is because in the event that neighbouring APs with operating channels bearing little or no frequency separation with each other, the effect of channel leakage, may be accounted during a MIPv6 or L2 handoff as reduced handoff delays.

D.6.2 Link contention and Router Advertisements

We have seen that Mobile IPv6 relies predominantly on the periodic transmission of Router advertisements to effect a fast movement detection and a subsequent IPv6 hand-

[10]On a number of experimental studies reporting significantly lower L2 or MIPv6 handoff times or latency optimisations [105, 410] we find little detail on the setup of the underlying channels over which neighbouring AP devices are operating. In such setup if APs are configured on neighbouring channels, an L2-handoff may experience better performance as a result of energy (and thus frames) leaking onto the neighbouring channel (i.e. being received from a neighbouring channel).

[11]It is important to distinguish between the channel leakage exploit used during the monitoring of the air interface by dedicated 802.11 monitor hosts and the avoidance of channel leakage at the MN by enforcing an inter-AP frequency separation of $> 25MHz$. We achieve this configuring APs to channel 2 and 9, while the WLAN monitors are configured at 4 and 8.

off. To this end, Mobile IPv6 requires a constant bit rate of router advertisement with a nominal average advertisement interval of 50ms; that is, Mobile IPv6 relies on near real-time guarantees of (periodic) delivery of this class of control signalling if the IPv6 handoff is expected to complete quickly.

While, timely (periodic) delivery of router advertisements works well for a single MN associated with the AP of the WLAN network, it is unclear if the such performance persists when more than one MNs associate and effect their communications over the same WLAN IPv6 network.

To assess the performance of router advertisements and the accuracy of the respective transmission interval we conduct a series of simulations whereby a constant bit rate flow simulating the router advertisement packets is sent by the AP over the air interface, while a number of wireless hosts associated that AP effect a non-saturating traffic mix of VoIP and HTTP communications with each other.

To magnify the effect of contention, without saturating the WLAN link we vary the size of the Beacon interval, identifying the magnitude of contention-free period repetition rate (CFPRate), also known as CFP repetition interval. CFPRate identifies the time period between which the PCF access method may[12] alternate in succession with DCF at the MAC sublayer of 802.11b, as shown in figure D.18

Supporting IPv6 flows with real or near-real time guarantees in IEEE 802.11 Wireless LANs, requires that the medium access control (MAC) layer provides similar guarantees for delivery of such MAC frames to participating MNs within a WLAN coverage area [411]; this is dependent on the number of wireless STAs that can be accommodated at a given bandwidth with specific delay bounds for wireless medium access [412].

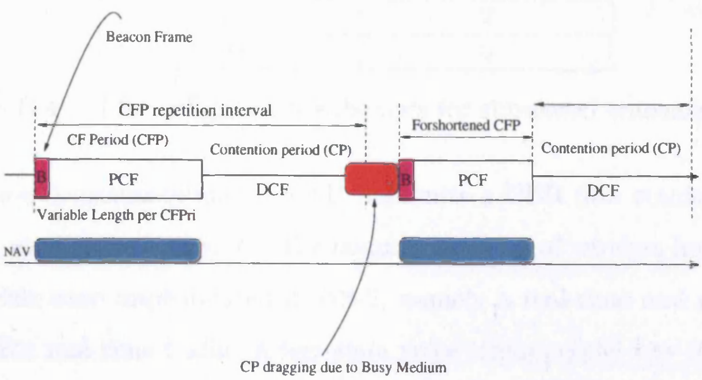


Figure D.18: Inducing contention (DCF) by varying CFPRate

[12] if implemented

Effectively, by varying the Beacon interval, both contention-free (CFP) and contention (CP) periods are affected; thus for a small number of MN associating with the AP, increased contention would reflect on the mean delay experienced by an 802.11 frame (MPDU[13]) prior its transmission.

Recent work by Suzuki & Tasaka [413], shows that if CFPRate is set too long (well above 300 ms), the guarantees for real-time delivery of IP traffic deteriorate drastically. Their results claim that for $CFPRate = 80$ ms a number of 18 STAs can be accommodated, transmitting a video IP flow with average MPDU delays between 100-130 ms.

Simulation Model

The scenario simulated, was implemented under the NS-2 simulator provided an implementation of both PCF [414] and DCF services. The simulation environment was extended by incorporating delay values with respect to the transmission of 802.11 frames, from our previous work in [51].

In this scenario, both PCF and DCF access methods were employed within a special node acting as AP in a simulated distribution system[14] (DS) comprising of a maximum of 10 STAs. Over PCF and for a period T_{cfp}, a set of ρ STAs transmitted real-time traffic, with the remainder $10 - \rho$ STAs transmitting, (over DCF and for time T_{cp}), non real-time data frames. Table D.4 provides the STA configuration sub-cases within the respective simulation scenarios.

No. of Voice STAs	No. of Http STAs
PCF	DCF
0	10
4	6
6	4
8	2
10	0

Table D.4: STA configuration sub-cases for simulated transmissions

During these communications the AP transmits a CBR flow emulating the router advertisement, sent every 50 ms. For the communications of wireless hosts, Two different traffic models were implemented in NS-2, namely a real-time and a non real-time traffic model. For real-time traffic, a *two-state* voice traffic model was considered [197]; During the ON-state (talk-spurt) it generates packets with fixed inter-arrival time ,

[13] MAC Protocol Data Unit

[14] a DS represents an 802.11 link operating at infrastructure mode, i.e requires the existence of an AP

while no packets are generated during the silence period (OFF-state).

Both states are distributed exponentially with mean for the ON-period $\lambda_{on} = 1000$ ms and OFF-period $\lambda_{off} = 1350$ ms in accordance with the well-established speech conversation model proposed by Brady [200]. During the ON-period the voice IP flow carried a payload of 24 bytes transmitted at 30 ms intervals. Such data rate specification is compliant with G.723.1 codec at 6.3 Kbps. An extra 40 bytes of headers above the MAC sublayer was also accounted (i.e IP+UDP+RTP).

In terms of non real-time traffic, a web traffic pattern was generated, by means of HTTP agents on STAs. HTTP client agents were located within some non-AP STA, while the server agent was assigned to the STA that acts as the AP. In a similar fashion, the HTTP client produced requests according to a two-state model, where the OFF-period was *Pareto*-distributed, while the ON-period followed a Weibull distribution [415]. The packet size for HTTP requests maintained a constant size of 250 bytes. Responses to HTTP requests were also Pareto distributed with $\alpha = 1.06$ and average response size $k_{avg} = 17.5$ Kbytes, while $k_{min} = 1$ Kbyte.

At the physical layer, the simulation considered a short physical layer convergence protocol (PLCP) preamble and header, resulting a maximum transmission cost of 96 μsec. Turnaround times and propagation delay were considered to be negligible since an error-free channel was assumed during all simulations. Three signalling rates were simulated, namely 2, 5.5 and 11 Mbps to establish discrete behaviour of dynamic rate shifting, as the STA moves towards the perimeter of the coverage area; in an IP mobility-enabled WLAN, this is where an IP handoff will be imminent.

Each simulation run had a duration of 1000 sec, while each variation of ρ as the number of CFP-active STAs was simulated for 40 iterations. Frame delay values were plotted against CFPRate at a 95% confidence interval of statistical significance. Table D.5 provides a summary of parameters configured during the simulations. T_{SIFS}, T_{DIFS} and T_{PIFS} denote the time periods for the respective types of interframe spacing (IFS). T_{slot} denotes the period of a single time slot. CW_{min} and CW_{max} denote the minimum and maximum contention window during DCF while `CFPRateMax` [17] denotes the maximum CFPRate considered in the simulation environment.

Simulation results

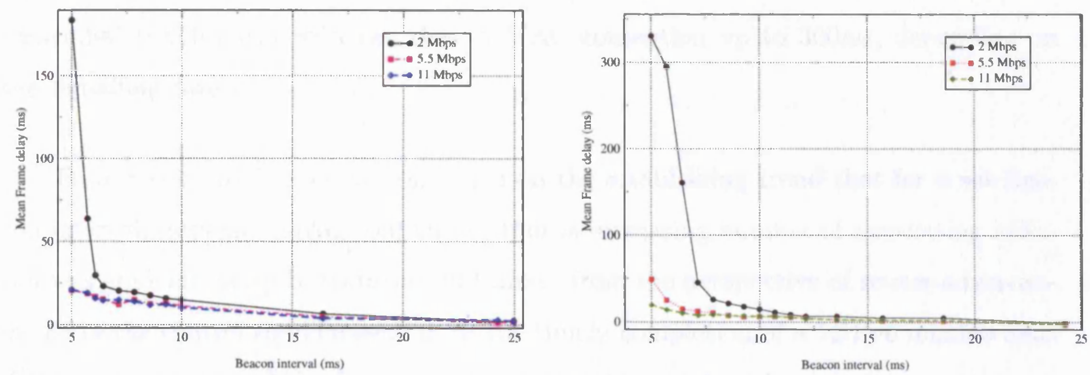
Figure D.19(a) shows the mean frame delay of router advertisements sent by the AP for 4 associated nodes. Frame delay is significantly higher for a low signalling rate as opposed to a high signalling rate. This is because the modulation scheme effected at

Parameter	Value
T_{SIFS}	10 μs
T_{slot}	20 μs
T_{PIFS}	$T_{SIFS} + T_{slot}$
T_{DIFS}	$T_{SIFS} + T_{slot}$
CW_{min}	31
CW_{max}	1023
CFPRateMax	30 ms
Voice λ_{on}	1000 ms
Voice λ_{off}	1350 ms
HTTP α	1.06
HTTP k_{avg}	17.5 Kbytes
HTTP k_{min}	1 Kbyte
CFPMaxDuration	[4,25]
No. of STAs	[4, 10]
Sig. Rate	2, 5.5, 11 Mbps
BER	0
PHY header	96 μs

Table D.5: MAC Layer Simulation Parameters

this signalling rate (QPSK), packs only two bits/symbol [51] transmitted.

By contrast, 5.5 and 11 Mbps employ a much faster modulation scheme (CCK), which can pack twice as many bits per symbol than the 2 Mbps rate can; while both 5.5 and 11 Mbps rates use the same modulation scheme, 5.5 Mbps rate packs 4 bits/symbol, while the 11 Mbps rate packs 8 bits/symbol.



(a) 4 nodes

(b) 6 nodes

Figure D.19: Mean frame delay of router advertisement flow for 4 and 6 associated nodes

The above imply that, the 5.5 and 11 Mbps rates would complete their symbol transmission sooner than the 2 Mbps rate. It can be seen from the graph of Figure D.19(a), that for small CFPRate values (around 6.2 ms), the 2 Mbps rate does not have sufficient time to complete frame transmissions from all 4 voice STAs (apparently exceeds CFPMaxDuration). Thus, polling of the STA list tail in PCF is postponed until CFPRate is renewed; but renewal of CFPRate is only effected at the end of T_{cp}, i.e. at the end of DCF. Hence, the excessive delay difference between 2 and 5.5 Mbps.

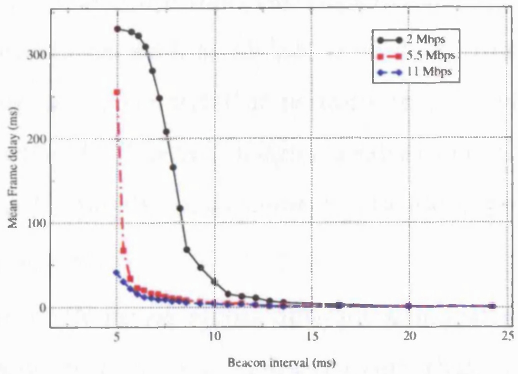
The observation is further verified as the number of voice STAs is increasing (figures D.19(b) and D.20(a). In particular, increasing the voice STAs by two, approximately doubles the frame delay at 2 Mbps for $CFPRate \leq 8$ ms with the respective router advertisement interval offset by that amount of delay shown in Figure D.19(b). While the peak frame delay value does not increase further as the number of voice STAs increases, it is reached sooner[15] by voice traffic sent at 5.5 and 11 Mbps, as shown in Figure D.20(b). In particular, the 2 Mbps rate, sustains high frame delays for larger $CFPRate \leq 11$ ms, while 5.5 and 11 Mbps approach delays in excess of 100 ms very quickly for $CFPRate \leq 7.5$ ms.

It can be seen, thus, that when the CFP repetition interval becomes small in relation to the number of associating MNs, the frame transmission delay increases significantly such that it effectively offsets the accuracy of isochrony in CBR transmissions. The importance stems from the fact that for small beacon renewal periods in relation to the number of associating MNs the MAC layer witnesses increased contention during busy (but not saturated) IP WLAN cell conditions; this becomes clearer from figure D.20(b); within a beacon interval of 6-10ms and for a small number of wireless STAs (ten) the transmission delay can increases due to MAC contention up to 300ms, depending on the signalling rate.

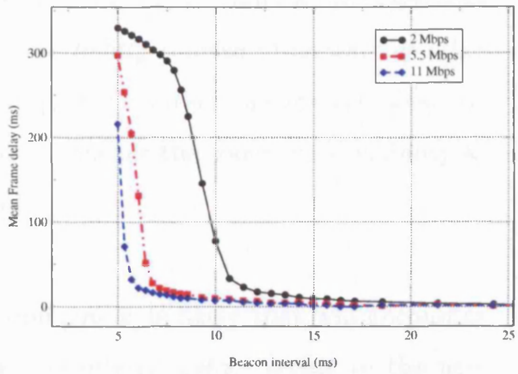
From the above it may be concluded on the establishing trend that for a set Beacon interval, increased MAC contention -that is increasing number of associating MNs-causes significant delay to transmitted frames; from the perspective of router advertisements as the control signal essential for the timely completion of a MIPv6 handoff such delay translates to a shift of the expected transmission interval.

For instance, in the event that the packet was scheduled for transmission every 50ms (router advertisement), due to increased contention the configured time interval would be offset by 250ms becoming an effective transmission interval of 300ms (Figure D.20(b)). It follows that, such interval shift becomes worse if the advertisement

[15] steeper climb of curve at peak values



(a) 8 nodes



(b) 10 nodes

Figure D.20: Mean frame delay of router advertisement flow for 8 and 10 associated nodes

interval is reduced; this is so because by scheduling more frequently advertisement frame transmissions, the MAC layer will experience increased contention, given that all MNs under 802.11x have an equal probability of capturing the air interface during DCF.

Thus, for an increasing number of associated nodes in a WLAN cell, the router advertisement interval is expected to *dilate* and as such, vary significantly the receipt time of this signal from an MN attaching the underlying visited AR; this is guaranteed to introduce significant latency during the MIPv6 handoff.

D.6.3 Operational Viability of MIPv6 in Internet Service Provisioning

While the Mobile IPv6 standard caters for a transparent mapping of addressing and routing state of the MN between previous and new points of IP attachment, it makes no explicit provisions for tight coupling of signalling context critical to IP connectivity. For instance, despite the fact that a transparent mapping - however fast, robust or reliable - allows the MN to roam between networks, there is no explicit coupling mechanism(s) to signalling provisions, by MIPv6, that would enforce access control of the MN; the same applies for providing Quality of Services assurance either through Integrated [416] or through Differentiated Services [417] for the MN that is IPv6-mobility capable.

Such coupling of signalling provisions are important not only as a transparent mapping of credential exchange or QoS setup, but also as a mapping that would provide - *at the same time - delay transparency* to the application layer. The latter has already been attested by previous elaboration on the sensitivity of interactive real-time services to delay overheads.

The above thus, two suggest once more the need for the IP mobility management standards, such as Mobile IPv6, to further *support tightly coupled provisions of other signalling context that pertains to IP connectivity*, while *preserving the seamlessness principle*. The two design considerations are paramount for the commercial viability as an IP mobility management networking protocol.

AAA State

If a MN moves across different administrative domains it is likely that will encounter some form of AAA infrastructure that must be negotiated before access to the new network can be granted. In the event of AAA establishment the set of interactions involved encompass a handshake between the MN, the local AAA server (AAAL) and the MN's home AAA server (AAAH). An attendant in the local network will ask for credentials from the MN and pass this on to the AAAL. The AAAL will then need to verify the identity of the MN with the MN's AAAH before it can grant access to the network. This implies that there is a requirement for at least one RTT between the AAAL and the AAAH to verify the MN and then another RTT between the AAAL and the MN to acknowledge verification. This may be reduced to a single RTT if the AAAL is co-located at the NAR. The important point is that one full RTT is required to authenticate and bill the MN. The size of this latency is dependent on the exact locations of the MN, AAAL, AAAH and the particular type of AAA implementation involved. Suffice to say, the incurred latency will be beyond that which is needed for the handoff to be considered seamless.

It is intuitive that a failure of AAA establishment due to lack of credentials, errors in the network, or errors with authentication protocols, is disastrous to the successfully completion of an IPv6 handoff, since access to the new network will be refused.

QoS State

Should the MN have particular QoS requirements it may need to convey this information to the new network. Using an Integrated Services/RSVP approach, the QoS signalling occurs separate to actual data transmission and therefore should not incur any additional latency for the handoff. In other words, no QoS state has to be established in order for the handoff to be successful. Of course, should required QoS parameters fail to be negotiated before the handoff completes, or the QoS requirements are rejected by the new network, then existing application sessions may suffer as a result.

A similar situation exists when using Differentiated Services (DS). No prior negotiation with any DS QoS broker and related policy servers need to have happened for the handoff to be successful and so handoff latency should not be affected. However, late successful negotiation of DS parameters or failure may harm existing application sessions that rely on certain levels of QoS in the network. One obvious example here is of using VoIP from a MIPv6-enabled WLAN device while performing a handoff between networks.

Contribution of other IP connectivity state to IPv6 handoff latency

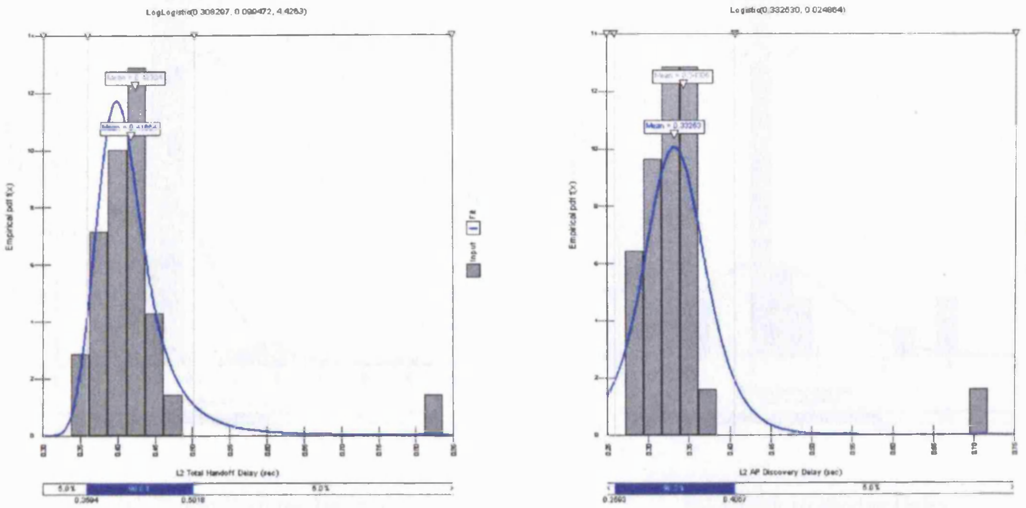
The time between the establishment of a globally routable care-of IPv6 address and the establishment of the appropriate context state defines an additional latency component contributing to the total delay of IPv6 handoff latency. Example contexts of such state pertaining to IP connectivity have been elaborated in previous sections for both authentication and billing as well as quality of service bounds.

We thus, introduce IP context establishment time, denoted by t_{ct}; it is defined as the delay required to establish/configure specific context pertaining to the IP connectivity context of the MN. Specific contexts of reference in IP mobility management in relation to Wireless LANs are:

The new total IPv6 handoff delay may incorporates also the t_{ct} component as follows:

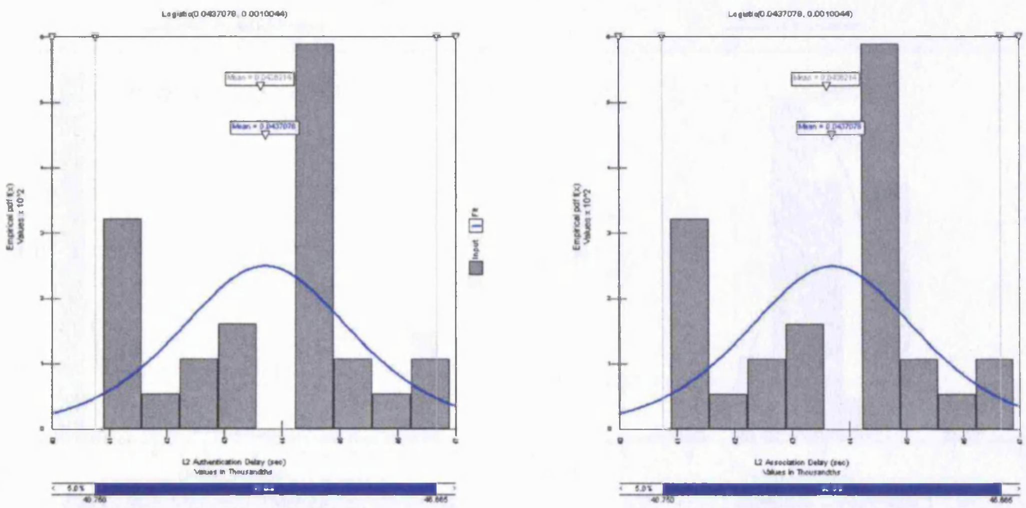
$$t_h = t_d + t_c + t_r + t_{ct} + t_o \tag{D.13}$$

D.7 Statistical Distributions of L2-handoff delay over 802.11



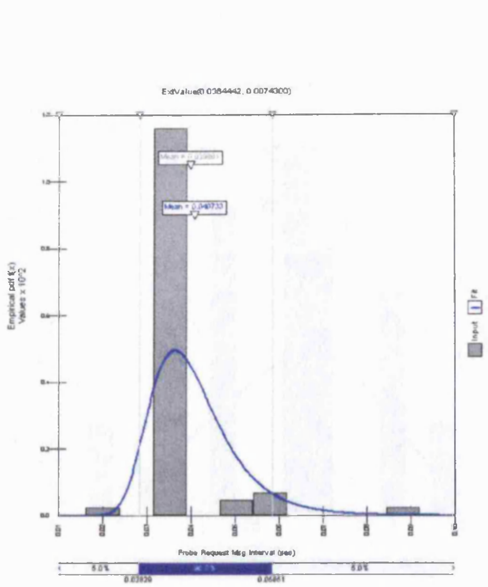
(a) Total L2-handoff delay (b) AP Discovery delay

Figure D.21: Distribution for total L2-handoff delay and AP-discovery delay component at 1 Mbps

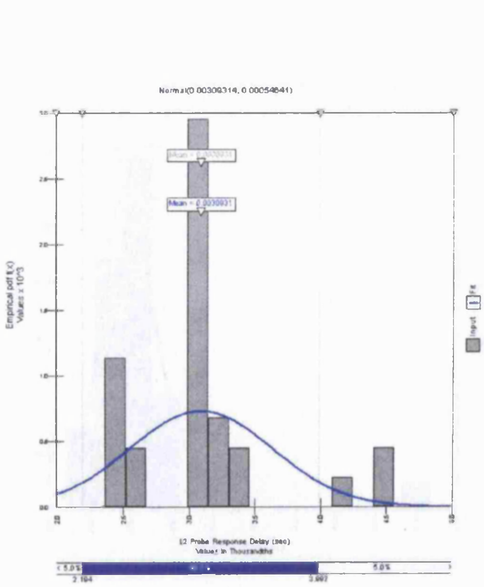


(a) Authentication delay (b) Association delay

Figure D.22: Distribution for Authentication delay and Association delay component at 1 Mbps

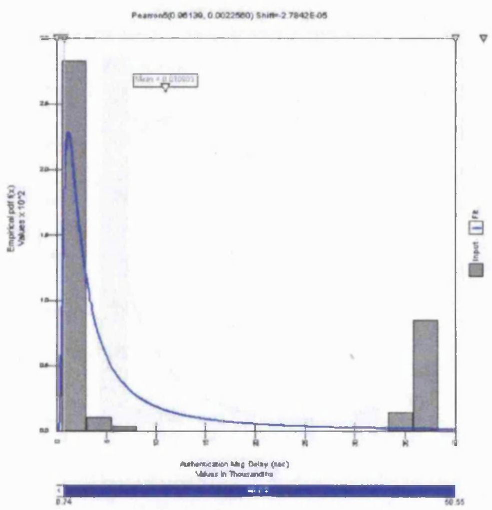


(a) Probe Request Interval

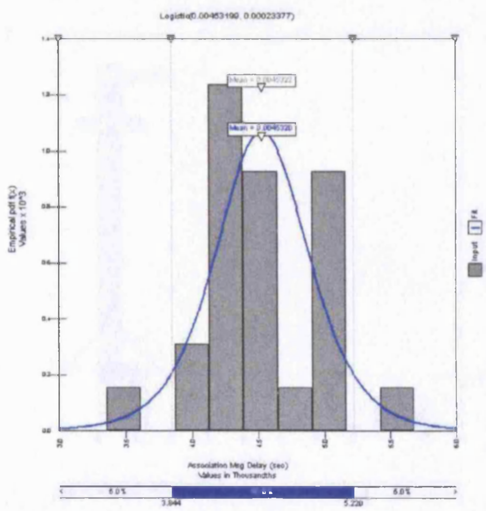


(b) Probe Response Delay

Figure D.23: Distribution for Probe Request (delay) interval and Probe Response delay at 1 Mbps

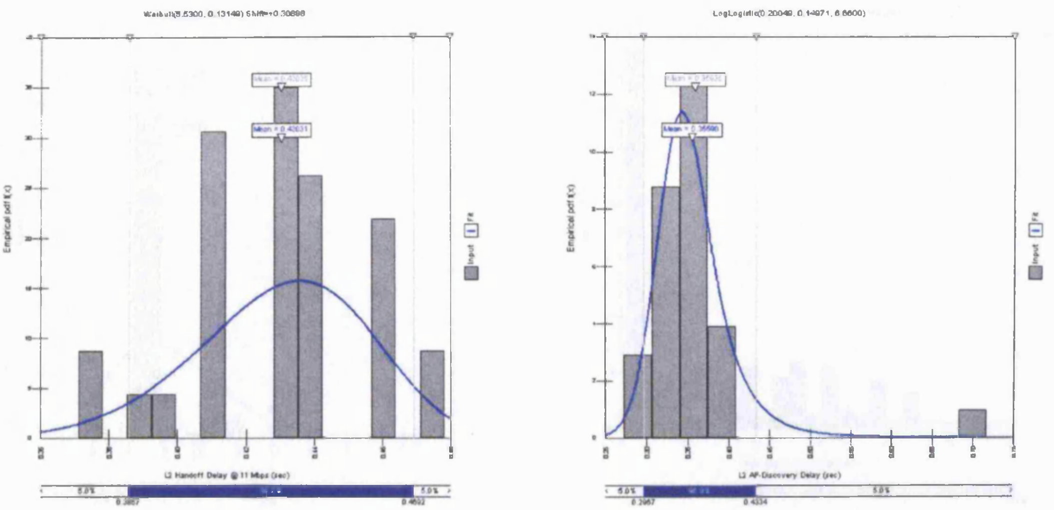


(a) Authentication Msg Delay



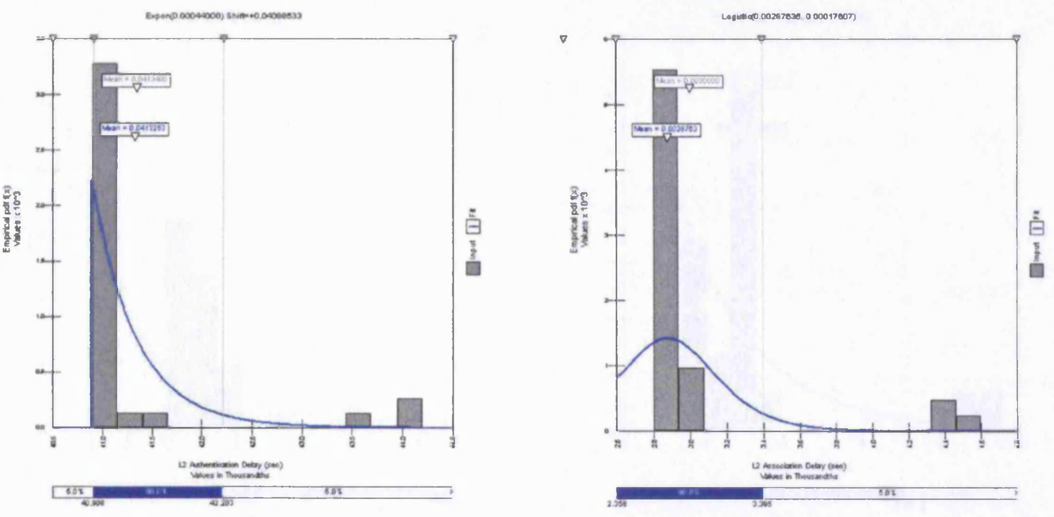
(b) Association Msg Delay

Figure D.24: Distribution for per-message Authentication delay and per-message Association delay at 1 Mbps



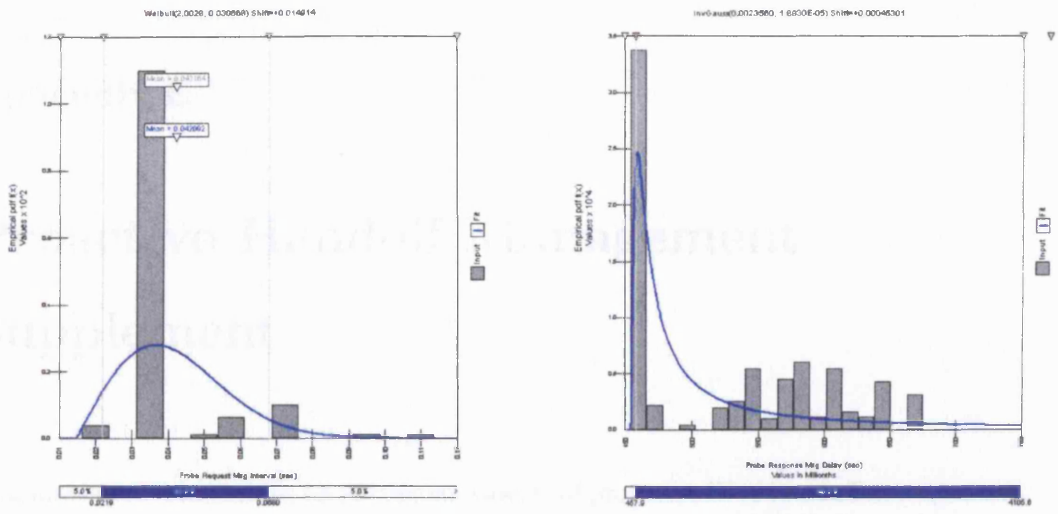
(a) Total L2-handoff delay (b) AP Discovery delay

Figure D.25: Distribution for total L2-handoff delay and AP-discovery delay component at 11 Mbps



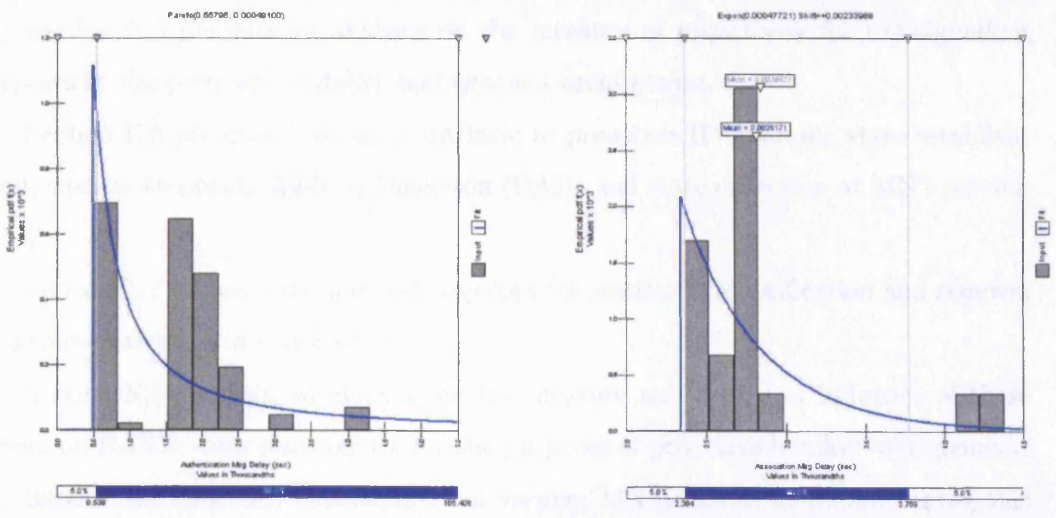
(a) Authentication delay (b) Association delay

Figure D.26: Distribution for Authentication delay and Association delay component at 1 Mbps



(a) Probe Request Interval (b) Probe Response Delay

Figure D.27: Distribution for Probe Request (delay) interval and Probe Response delay at 1 Mbps



(a) Authentication Msg Delay (b) Association Msg Delay

Figure D.28: Distribution for per-message Authentication delay and per-message Association delay at 1 Mbps

Appendix E

Proactive Handoff Management Supplement

This annex provides details on particular aspects of proactive IPv6 handoff management.

In particular, Section E.1 presents state extensions relevant to a differentiation in the type of propagation environment, together with bootstrap learning dynamic for HAR discovery (Section E.2).

Section E.3 provides a detailed description of both direct and indirect HAR updates together with minor optimisations to boost HARD state convergence.

Section E.4 presents a mechanism to support robust resolution of malicious HAR hints during HAR discovery.

Section E.5 presents an analysis on the measure of proactivity for the signalling purposes in the proposed mobility management architecture.

Section E.6 presents operations intrinsic to proactive IP Roaming state establishment, such as Duplicate Address Detection (DAD) and state collection at MN's serving AR.

Section E.7 presents the methods employs for statistical identification and removal of extreme values from collected traces.

Section E.8 presents an analysis on the measure and degree of influence of these factors on HARD state convergence for the purposes of proactive handoff management.

Section E.9 provides visualisations of varying MN densities employed during this simulation study for a nominal network topology of 45 PoAs. Section E.10 presents the corresponding density of random way-points for different MN populations sizes

Section E.11 provide the plotted confidence intervals for Handoff Performance under varying pause movement periods.

Section E.12 provides the plotted measure of service utility emerging as a result of proactive handoff management for larger variations on the density of PoA topology. These are accompanied by the respective probability density functions for these PoA densities.

Additionally it augments the set of experimental results on HARD and proactive handoff performance coupled with more complex PoA topologies employed during the course of the simulations.

E.1 Differentiation of Indoor/Outdoor coverage for an M-neighbourhood

For the purposes of optimising the set of handoff PoA neighbours, that are candidates for MN's next IP handoff, different predictive handoff techniques may be employed depending on the type of the propagation environment [418, 419, 420]. To this end, the M-neighbourhood identification process may distinguish between two forms of wireless coverage: *indoor* and *outdoor*.

For the purposes of *outdoor* coverage, with little or no obstructions within the serving CA, its respective AR, may support a pre-configured *Coverage Area Tuple (CAT)*, encompassing AP's geodetic coordinates. This is denoted as:

$$CAT_{APID_i} = \{(l_i, L_i, r_i) : APID_i \ \epsilon \ MNV_{k,i}\} \tag{E.1}$$

where l_i the latitude position of $APID_i$, (L_i) it longitude and r_i its radius. Range information is typically available for AP wireless technologies in the form similar to the one of table E.1. Location information is assumed to be available during installation of the AP in a similar fashion as in [421].

Depending on whether the AR is configured for indoor or outdoor IP handoff management, a CAT tuple may, thus, augment the MNV-element of each PoA:

$$MNVe_{k,i} = (APID_i, Channel_{APIDi}, \theta_{APIDi}, DomainID_l, CAT_{APID_i}) \tag{E.2}$$

For indoor coverage the MNV-RNV mapping may not support location coordinates, since the distance between APs as well as the size of their coverage areas is expected to be small. Table E.1, presents a typical transmission range specification for indoors and outdoors deployment of IEEE802.11b WLAN APs.

It is also noted that closed-space propagation environments make impractical[1] to attain latitude/longitude position coordinates. Thus, for indoors wireless access networks the respective CAT tuple has a *null* value.

propagation \ B/w (Mbps)	11	5.5	2	1
Obstructed (indoors) range (m)	25	35	40	50
Semi-obstructed (in/outdoors) range (m)	50	70	90	115
Free Space (outdoors) range (m)	160	270	400	550

Table E.1: Transmission range specification of IEEE802.11b

E.2 Bootstrap learning dynamics for Handoff AR discovery

Typically, the topology of the wire-line segment of fixed infrastructure wireless networks does not change often. The same is the case for wireless technologies implementing APs within licensed frequency bands.

However, for deregulated wireless technologies operating in the unlicensed industrial, scientific, manufacturing (ISM) frequency bands of 2.4 and 5GHz, AP implementations allow fast deployment and expansion of the wireless access segment. As a result, new PoAs comprising of single AR-AP systems may augment the access network infrastructure. Alternatively, providers may expand bandwidth capacity or wireless coverage by simply increasing the number of serving AP within a access network domain.

In such cases it appears increasing difficult to initialise *manually* each member of a joint M-R-neighbourhood, with operational APs and associated ARs. This requires a *bootstrap* process of initial *MNV-RNV* mapping discovery supported by a *dynamic learning and configuration* mechanism.

Such form of dynamic learning, however, may be supported by information conveyed proactively to AR members of the R-neighbourhood, by the MN. Handoff AR (HAR) identification may, thus, exploit the temporal existence of bypassing MNs.

Dynamic learning of this type fits naturally to MN movement; bypassing MNs discover explicitly the identity of associated APs, as well as the address of the serving AR, when associating with each point of attachment (PoA), while in transit towards a destination. As a result the MN can provide two types of information: (i) the identity of the associated AP_i during attachment to the controlling AR_i; (ii) the identity of

[1]position coordinate implementation aids such as Global positioning system (GPS) cannot operate in closed-space environments

both AP and AR $(APID_i, AR_i)$ at the previous PoA_i during attachment with the new PoA_{i+1}.

AP identity information is usually available to the link-layer of the MN. the AP identifier is essential for the MN to differentiate between APs that are detected during the link-layer handoff process. For certain wireless access technologies this is essential to allow multiple APs to operate on the *same* frequency channel. The identity of the AP provides a differentiator for the purposes of selection amongst the APs available during the scanning stage of an L2 handoff. The AP identity is also used by the MN for the purposes of association with an AP, since certain technologies allow association only with a single AP; this is the case for instance with 802.11 WLAN technology.

It should be noted that for certain technologies AP identity information is typically available to the user for access diversity purposes. This implies that link-layer information can be available to other layers of the network stack. As such it can also be available to the network layer for handoff management purposes. From a bootstrap perspective, AP identity information allows the AR to identify all APs directly attached to it, with no need for manual configuration from the network administrator. In this manner, the initial MNV-RNV mapping between an AR and its locally attached APs can be automated and remain accurate independent of the deployment requirements of each wireless access network.

HAR discovery exploits the availability of AP identity information at the MN to initialise the MNV-RNV mapping at the attached AR. It requires the MN to communicate proactively to the associated AR, the identity of the AP currently associated with the MN. Beyond the bootstrap stage, HAR discovery must identify the handoff AR neighbours to MN's current AR.

E.3 Handoff AR discovery algorithms

E.3.1 Indirect (Hinted) Handoff AR updates

Prior to formation of a joint M-R-neighbourhood, each AR bootstraps HAR discovery by initialising its own MNV-RNV element from the local AP information provided by the MN. When the MN establishes an association with the first PoA, its current AR (AR_c), conveys its RNV-element onto the MN comprises of the information identified in Section 4.7.1.

Upon completion of its next IPv6 handoff to the new AR (AR_n), the MN receives a router advertisement, announcing the IPv6 address of the AR serving that network link. The MN checks the prefix against a small cache of *visited PoAs*, to ensure that it resides on a different network, compared to past IP-cell transitions. As soon as it verifies that a new IPv6 network has been encountered, it sends an *indirect RNV update* to AR_n as shown in figure E.1(a). The new AR updates its corresponding MNV and RNV vectors stored in an *Mroute Cache*, while the entry's flag is set to `TENTATIVE`. The incoming RNV-element cannot be used until an *acceptance threshold* $Mroute_{accept}$ of q identical RNV updates have been reached, by an equal amount of *different* MN arriving in AR_n. Assuming a unique MN identity $MNid_i$ such measure confirms the validity of the received RNV-element confirming HAR neighbourhood of MN's previous AR (AR_p).

When the Mroute acceptance threshold has been reached, the AR_n can remove the restriction of the tentative RNV-element, by changing its status to `ACTIVE`. Entries that do not increase their acceptance threshold after q RNV updates received[2] return back to their `TENTATIVE` state; as such they cannot be used in the buildup or propagation of MNV-RNV mappings to other HAR neighbours. Entries that remain in the `TENTATIVE` for more than q bypassing MNs (handoffs), are *removed* from the Mroute cache. The algorithm is illustrated in Algorithm 1.

For optimisation purposes, an indirect RNV update may be conveyed to AR_c as an option of a Neighbour Advertisement sent by the MN towards AR_c. A similar practice may be followed when AR_c conveys its own RNV element back to the MN. In this manner, the HARD mechanism refrains from generating additional signalling. Hinted-HAR updates allow new PoAs to identify AR neighbours with no inter-AR communication. As a result, the efficiency of such mechanism is highly dependent on the rate of bypassing MN from a particular PoA.

E.3.2 Charged Handoff AR updates

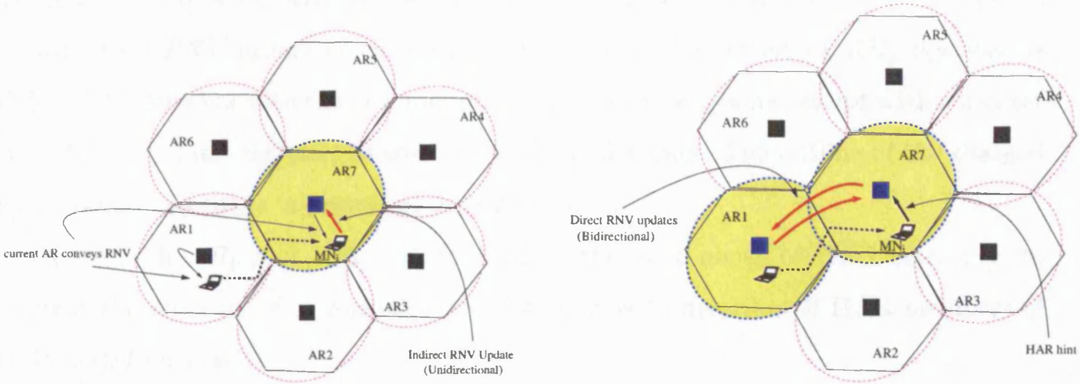
A charged HAR update places less emphasis on the content of the AR neighbourhood information hinted by the MN, while focusing on inter-AR communication for the purposes of deriving Handoff AR neighbourhood characteristics.

The underlying algorithm can be viewed from two perspectives depending on the implementation: (i) a hinted HAR update is extended such that the new PoA can also propagate its RNV to the previous PoA; (ii) an MN hint that carries no RNV information except for the IPv6 address of AR_p at the previous PoA and any local AP identifier that may initialise the MNV of the new PoA. Such hint acts only as a trigger

[2]or alternatively a fixed period t_{revoke}

Algorithm 1 Indirect RNV update through MN hints

Require: $MNid_i$ is unique
Ensure: MN_i interface is up
  while MN_i is roaming **do**
    AR_{i-1} send $Mroute_c$ to MN_i
    $Mroute_c$ overwrites any previous Mroute entry MN_i
        ⋮
    MN_i handoff to AR_n
    % AR_c transits to 'previous' state
    $AR_p \Leftarrow AR_c$
    if $ARprefix_{RtAdv} \notin PoA_{cache}$ **then**
      MN_i send $(Mroute_c, AP_{current})$ to AR_n
      %Augment bootstrap MNV AR_n ————-
      Append$(MNVe, AP_{current})$
      %Compute m-routing at AR_n ————-
      $Mroute_{accept} = q_{updates}$
      if $Mroute_c \in Mroute_{table}[j]$ **then**
        if $Mroute_{MNid} <> SrcAddr_{MN}$ **then**
          r++
          $Mroute_{flag} \Leftarrow ACTIVE$
          Update$(Mroute_{valid}[r], Mroute_c)$
        end if
      else
        j++
        $Mroute_{flag} \Leftarrow TENTATIVE$
        $Mroute_{wait} = Mroute_{accept}$
        $MN_{id} \Leftarrow SrcAddr_{MN}$
        $Mroute_{check}[j] = (Mroute_c, Mroute_{flag}, Mroute_{wait}, Mroute_{MNid})$
      end if
      ⋮
    else
      Ignore(RtAdv)
    end if
    % Remove stale 'TENTATIVE' HAR hints AR_n ————
    for k = 1 to j **do**
      if $(Mroute_{flag}[k] == TENTATIVE) \&\& (Expired(Mroute_{wait}[k]))$ **then**
        Delete$(Mroute_{check}[k], Mroute_c[k])$
      else
        $Mroute_{wait}[k]++$
      end if
    end for
    % New AR transits to Current state ————
    $AR_c \Leftarrow AR_n$
  end while



(a) Indirect RNV HAR update (b) Direct RNV HAR update

Figure E.1: Handoff AR discovery as a result of indirect or direct RNV updates hinted by the MN

for the new PoA to initiate a handoff AR identification signalling exchange, where both AR_n and AR_p provider their MNV-RNV mapping to each other.

In both cases, the difference amounts to the level of *trust* put on an MN; in the first case the MN is trusted to convey RNV-MNV mappings between two PoAs; in the second the MN cannot be trusted with the content of RNV-MNV mappings between two PoAs for reasons of infrastructural security. The network does not want to divulge wireless transmission capabilities of its APs or their exact coordinate location when a CAT tuple is supported. However, the amount of signalling remains the same in both implementation scenarios. To simplify description, the second implementation scenario is assumed.

Under the charged-HAR update algorithm, the current AR (AR_c) does *not* convey any MNV-RNV elements to the MN. Instead, the MN simply stores the IPv6 address of AR_c within its *visited PoA* cache, identified from periodic router advertisements received at that network link.

As soon as the MN completes its handoff to the new AR (AR_n), it receives a router advertisement, announcing the AR's IPv6 address. The MN checks such address against the cache of visited PoAs, to confirm a new PoA and hints with the address of the *previous AR* the new AR (AR_n). AR_n stores the AR_p address hint in its *Handoff AR neighbours* table. In a manner similar to indirect RNV updates, AR_n refrains from using such information immediately until an acceptance threshold on the number of HAR-hints about the same AR_p have been received. Once the acceptance threshold at AR_n is satisfied the HAR hint is pursued to establish HAR neighbourhood information between the two ARs; AR_n sends its own RNV-MNV elements to AR_p, through a unicast *direct RNV update* as shown in figure E.1(b). The receiving AR_p updates its RNV-MNV elements with the new mapping, and acknowledges its receipt with a unicast direct RNV update, carrying its own RNV-MNV elements. The outline of the charged HAR update algorithm is presented in Algorithm 2.

Now, both AR_p and AR_n can proceed in the next phase of HAR discovery, to augment the measure of m-routing state that begins to describe all HAR members of the R-neighbourhood.

E.3.3 Boosting m-routing state convergence between Handoff ARs

In the previous sections, both indirect as well as direct RNV update algorithms have described a means of exchanging RNV-MNV mapping between the previous and the new PoA of a MN. In an effort to boost convergence of m-routing state information between *all* members of a handoff AR neighbourhood, the HAR discovery algorithm introduces a third HAR identification step described as *HAR discovery boost* (HARD-boost).

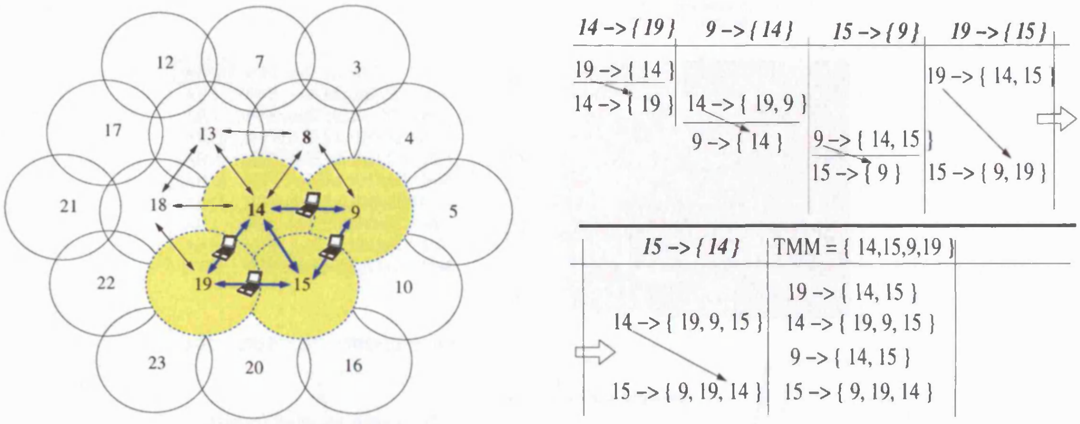
A HARD boost supplements, essentially, the step of a charged HAR update. Algorithm 2 prescribes that when MN's new PoA decides to pursue a verified HAR neighbourhood hint, it contacts the identified AR neighbour, to exchange R-neighbourhood information, through a direct RNV update. In the event that such direct RNV update is *boosted* to expedite m-route convergence, the dispatched direct RNV update includes two types of RNV-MNV mappings: (i) a *single* RNV-MNV mapping describing the particular AR_n, (ii) *all* established RNV-MNV mappings, verified by previous direct HAR neighbour exchanges.

In the above manner, during a single direct RNV update between each pair of HAR neighbours, can achieve fast convergence of m-routing state for a small number of MNs transiting within the R-neighbourhood of the serving PoA. Figure E.2(a) illustrates how a partial HAR neighbourhood of four ARs begins to form as a result of MN movement. Figure E.2(b) presents the steps of the algorithm. For the purposes of description, each step is described sequentially. However, under operational conditions these steps occur asynchronously and thus potentially in parallel.

The first movement emanates from PoA_{14} towards PoA_{19}. The underlying ARs have no existing HAR state and as such, exchange only their bootstrap RNV vectors. The second movement arising from PoA_9 towards PoA_{14} causes an RNV exchange including HAR state pertaining to PoA_{19}. After five different MN movements between four neighbouring PoAs, the *Tentative Mobility Matrix* (TMM) describing their HAR

Algorithm 2 'Charged' (Hint+Exchange) Handoff AR updates

Require: $MNid_i$ is unique
Ensure: MN_i interface is up
  % At MN
  while MN_i is roaming **do**
    MN_i handoff to AR_i
    % Current AR transits to Previous state
    $AR_p \Leftarrow AR_c$
    MN_i receives RtAdv
    if $ARprefix_n \notin PoA_{cache}$ **then**
      MN_i send $IPaddr_ARp$ to new AR_i
        ⋮
      Append($MNVe, AP_{current}$)
      % Compute m-routing at AR_n ——————-
      $Mroute_{accept} = q_{updates}$
      if $IPaddr_ARp \in CAR_{table}[j]$ **then**
        if $MN_{id} <> SrcAddr_{MN}$ **then**
          $SrcAddr_{MN} \Leftarrow ACTIVE$
          AR_n send $Mroute_n$ to AR_p
            ⋮
          % At AR_p ——————————
          if $Mroute_n \notin Mroute_{table}$ **then**
            r++
            Update($Mroute_{table}[r], Mroute_n$)
          else
            Discard($Mroute_n$)
          end if
          AR_p responds with RNV-MNV element to AR_n
            ⋮
          % At AR_n ——————————
          if $Mroute_p \notin Mroute_{table}$ **then**
            Update($Mroute_{table}, Mroute_p$)
          else
            Discard($Mroute_p$)
          end if
        else
          Discard($IPaddr_ARp$)
        end if
        ⋮
      else
        j++
        $SrcAddr_{MN} \Leftarrow TENTATIVE$
        $MN_{id} \Leftarrow SrcAddr_{MN}$
        $CAR_{table}[j] = (IPaddr_ARp, MN_{id})$
      end if
      Update($ARprefix_n, PoA_{cache}$)
    else
      Ignore(RtAdv)
    end if
    % Remove stale 'TENTATIVE' HAR hints AR_n –[see Algorithm 1]–
    % New AR transits to Current state ——————
    $AR_c \Leftarrow AR_n$
  end while



(a) MN movement and HAR update sequence (b) HAR boost algorithm

Figure E.2: Boosting Handoff AR discovery with RNV vectors from all HAR neighbours per direct RNV update

neighbourhood is complete in all AR neighbours.

The maximum number of steps required by the HARD algorithm to produce a complete tentative mobility matrix in *all* members of a HAR neighbourhood, is dependent on the size of the HAR neighbourhood. A HAR neighbourhood can be represented as an undirected wheel graph of order n with each vertex representing a single PoA [422]. A wheel graph contains a cycle graph of order $n-1$ and a trivial complete graph of one vertex identified as the hub. The number of edges connecting the wheel's vertices are $2n$.

This implies a total of $4n$ possible mobility paths available to the MN in any direction between adjacent AR within a HAR neighbourhood. Boosted HAR updates allows any AR member of the HAR neighbourhood to obtain a complete TMM within $2n$ MN movements in any different direction within that neighbourhood such that each PoA is visited only once. The complete TMM at the current AR of an MN is shown in figure E.3.

Convergence of m-routing state provides robust support of *IP-mobility-hop* routing within that HAR neighbourhood. This is also essential for seamless flow forwarding management purposes presented in Chapter 5.

In addition, knowledge of HAR neighbourhood information allows the serving AR to make advance planning about MN's next IPv6 handoff. The form in which, proactive planning of MN's next IPv6 handoff must be pursued, is dependent now on the type of IP connectivity state required by the handoff process.

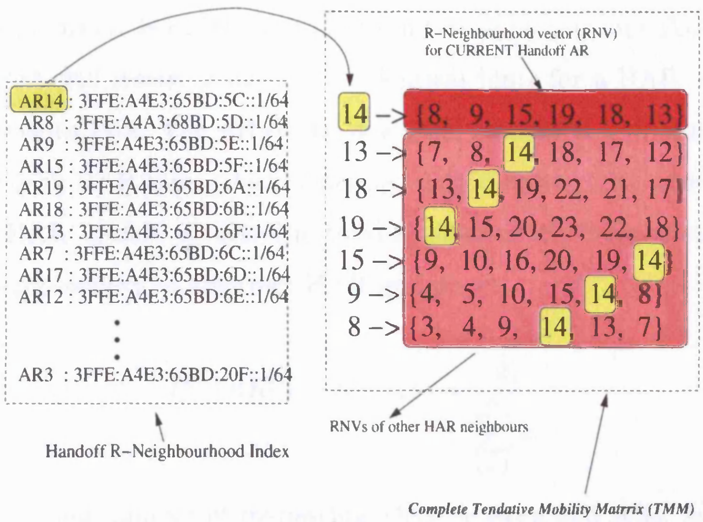


Figure E.3: m-routing state (tentative mobility matrix) maintained in PoA_{14} for its underlying HAR neighbourhood

E.4 Identifying malicious MN hints

To identify misbehaving MNs, it is first assumed that AR entities within a network domain operate securely. Misbehaving MN's in the context of HAR discovery may take two main forms in terms of erroneous MN HAR hint: (i) mistaken PoA information that is not adjacent to the previous PoA (ii) malicious PoA information targeting Denial of Service attacks for targeting either existing MNs or the serving AR.

For the first class of erroneous HAR hints provided by the MN, the problem does not amount to a malicious MN, but rather to erroneous protocol implementation at the MN; that is the MN does not attempt to maliciously misrepresent itself with different CoA identifiers at the new PoA.

In this case, erroneous HAR hints can be quickly identified by applying basic *swarm intelligence* [423] on the validity of such mappings. Swarm intelligence (SI) is the property of a system whereby the collective behaviours of (unsophisticated) agents that interact locally with their environment, cause coherent functional global patterns to emerge. Such emergent behaviour has been observed in biological organisms such as bacteria or insects.

For the purposes of HAR discovery based on MN hints, the new AR_n at the new PoA_{i+1} can apply differential swarm intelligence extracting the *erroneous* MN hint from the correct ones. In particular, as a number of MNs are moving from one PoA to another, each MN hints a HAR update to the new PoA identifying the previous PoA that the MN came from. For MNs arriving from the *same* previous PoA, it is intuitive that the new PoA will receive a number of identical hints for a HAR update pointing to the same previous PoA. The AR of the new PoA maintains a running *probability of correctness* for each HAR hint, which is defined as the ratio of the number of hints for the particular HAR update $\bar{x}_i$ over the total number of HAR hints observed at that PoA. Thus the probability of a correct HAR update is:

$$P(HARU_i = correct) = \frac{\bar{x}_i}{\sum_{i=1}^{n} \bar{k}_i} \tag{E.3}$$

For an increasing number of by-passing MNs between two ARs, false RNV-MNV mappings emerge clearly as *singularities* and hence, can be eliminated.

In the case of malicious MN HAR hints additional control is required primarily as a result of robust identification of the MN:

- the MN can misrepresent itself to the new PoA by attaining multiple new CoAs. It is important to note that under IPv6 *cannot* spoof the IPv6 address of another host, since during Neighbour Discovery the Duplicate Address detection will detect the duplicate and thus, will deny usage of this address to the malicious MN. Thus, attacks are not expected as a result of spoofed IPv6 CoA addresses.
- the MN can reconfigure its MAC-layer address
- the MN can attain multiple public-private keys if the association is dependent on the MAC identifier of the network interface.
- the MN can temporally change its Home address.
- the MN may identify itself with a bogus Home Agent

Under such such conditions and assuming that the home network of the MN can be trusted, the MN must be subject to remote authorisation and authentication by the local AAA authority of its home domain, while a Network Access Identifier (NAI) must be required [424]. That implies that the authenticity as well as authorisation of MN's identity can only be decided on the basis of validating MN's credentials with the local AAA authority of MN's home domain. This is the case for cellular systems, whereby each mobile terminal handset is assigned a unique International Module Subscriber Identifier (IMSI) permanently associated with the mobile subscriber identity module

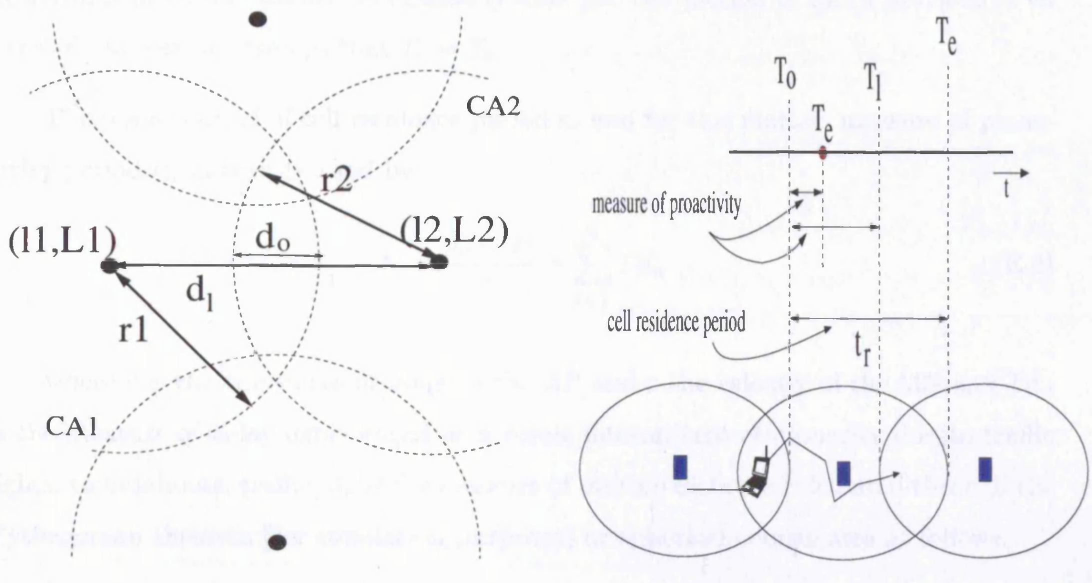
(SIM) card [425, 426]. For bootstrap purposes, the AR may check the AP within the HAR update provided by the MN, whether it belongs to the list of authorised APs within that domain [282].

The following sections present the HAR discovery algorithms based on hinted-HAR or HAR-charged updates. For the scope of this study, these algorithms are limited to erroneous HAR updates as a result of misconfiguration or erroneous implementation. For the purposes of malicious MN hints these algorithms must be extended to include AAA authentication and authorisation checks, which is beyond the scope of this study.

E.5 Measure of proactivity

Proactivity in state establishment, is an abstract time-based intervention with respect to the act of an event; in the case of mobility management, such event is MN's IPv6 handoff. However, by itself proactivity does not define the exact timing that state establishment must be initiated. It is, thus, important to address how early should context state be established or relocated before MN's next IPv6 handoff.

Essentially, the measure of proactivity is tracked by MN's *cell residence* period; this is defined as the time period between the time instant of MN's previous IPv6 handoff completion (T_o) and the initiation of its next IPv6 handoff (T_n), as shown in figure E.4(b). T_l represents the latest period where proactive state establishment must be pursued for enhanced IPv6 handoff management.



(a) Model of overlap distance

(b) time measure of proactivity

Figure E.4: Measure of proactivity as a function of cell residence period

For handoff management purposes it is imperative that the time instant of state establishment is located as closely as possible to the completion time of MN's previous handoff (T_o); that is, state establishment has to be pursued at the earliest time deadline, T_e. The reason is intuitive: if delay seamlessness is to be preserved, the MN must ensure that critical IP connectivity state for its next IPv6 handoff is pre-configured before the end of its cell residence period. Where proactivity acts in preparation of MN's next IPv6 handoff, the cell residence period remains unknown to the MN. As such, pre-configuration of IP connectivity state becomes a critical function with immediate deadline. This is particularly the case for *vertical* handoffs.

On the contrary, when proactivity acts in anticipation of MN's next IPv6 handoff, predictive techniques can provide some measure of its cell residence period. This can be exploited to delay proactive state establishment until absolutely necessary. Such observation is of particular significance for highly volatile capability information conveyed on-demand to the MN. If handoff bandwidth information capabilities are conveyed to the MN on-demand, then the MN must ensure that the request of such capabilities is made when close to the end of its residence period; this will prevent the MN enforcing a handoff decision based on inaccurate or time-variant[3] information.

For the purposes of IP Roaming state establishment, the information provided by the HAR neighbours is *persistent*. For this reason, the proposed model allows state establishment to be initiated immediately after the completion of MN's previous IPv6 handoff; namely, it assumes that $T_o = T_e$.

The upper bound of cell residence period t_r and for this matter, measure of proactivity period is, in turn tracked by:

$$t_r = \frac{d_o - 2r}{v} + \sum_{i=1}^{n} Td_n \tag{E.4}$$

where r is the transmission range of the AP and v the velocity of the MN and Td_n is the measure of delay experienced as a result intermittent stationarity due to traffic lights, roundabouts, traffic; d_o is the measure of overlap distance calculated through the Pythagorean theorem (for simulation purposes) or spherical coordinates as follows:

[3]The HAR neighbour producing such information may want to attach an expiry time based on time-of day statistics.

$$\begin{aligned} k_i &= d_l - r_i \\ k_{i+1} &= d_l - r_{i+1} \\ d_o &= d_l - (k_i + k_{i+1}) \end{aligned} \tag{E.5}$$

The distance between the APs[4], shown in Figure E.4(a), may be expressed as:

$$d_l = R \times d_a \ : \ R \ = 6.37123 \times 10^6 \ m \tag{E.6}$$

$$d_a = 2asin(min(1, \sqrt{sin(\frac{\delta l}{2})^2 + cos(l_i)cos(l_{i+1})sin(\frac{\delta L}{2})^2})) \tag{E.7}$$

$$where, \ \delta L = L_{i+1} - L_i \ \wedge \ \delta l = l_{i+1} - l_i$$

Assuming spherical[5] earth shape, d_a has been optimised by means of half angles to effect higher accuracy for small distances between APs. The value of d_o signifies the following:

$$d_o = \begin{cases} > 0 & , (APID_i, \ APID_{i+1}) \text{ adjacent} \\ = 0 & , (APID_i, \ APID_{i+1}) \text{ adjacent, but no overlap} \\ < 0 & , (APID_i, \ APID_{i+1}) \text{ non-adjacent} \end{cases} \tag{E.8}$$

Results show that equations (E.6) and (E.7) provide reasonable accuracy for adjacency purposes for CA radius above 100m, while reasonable overlap measures can be provided for CA radius above 300m.

E.5.1 Signalling overheads and state establishment optimisations

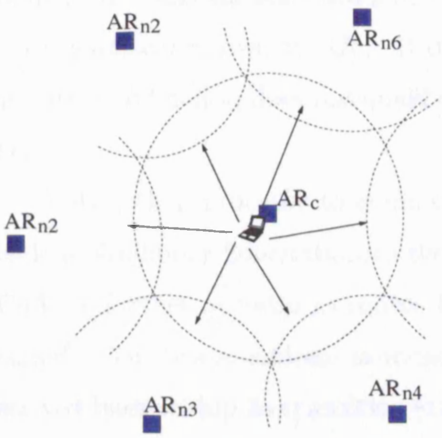
From the perspective of signalling overheads, the above mechanism may be identified as a *pessimistic state establishment approach* (see figure 4.7(a)); it requires no prediction of the exact next IPv6 handoff AR candidate at the cost of additional HAR signalling.

Under pessimistic state establishment, all members of the HAR neighbourhood have the same probability of being selected as the next PoA as shown in figure E.5(a). This is the case when the *short-term* direction within MN's movement remains highly irregular. In such case, *all* HAR neighbours need to establish state for that particular context requested by the MN ahead of its upcoming IPv6 handoff.

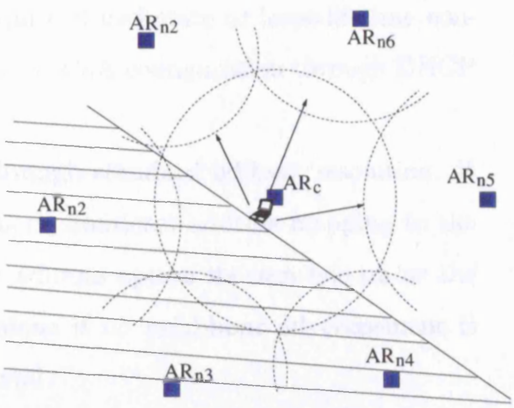
[4]centre of two adjacent CAs.

[5]Further distance optimisations may be effected by considering an ellipsoid Earth shape. However, this involves somewhat more complex calculations.

To reduce the amount of signalling overheads between HAR neighbours, an optimistic state establishment approach may be pursued. Under such scheme, handoff management takes the optimistic approach that certain AR neighbours emerge as low probability candidates for MN's next IPv6 handoff; this can be the case for regular (e.g. straight line) mobility patterns typically expressed through forward movement, as illustrated to figure E.5(b). Anticipatory proactive handoff management through optimistic forms of state establishment is beyond the scope of this study and thus proposed as future research direction.



(a) Pessimistic Approach



(b) Optimistic Approach

Figure E.5: Pessimistic or Optimistic perspective in state establishment within MN's HAR neighbourhood

This study focuses on the description and performance of the pessimistic state establishment approach for the purposes of IP roaming state establishment.

E.6 IP-Roaming state intrinsics

E.6.1 DAD during proactive state establishment

Duplicate Address Detection of an IPv6 sCoA is essentially address resolution for a tentative address [110]. This is usually performed link-locally upon generation of the new IPv6 address by the host. If a neighbour solicitation for that IPv6 address is not responded with an advertisement within `RetransTime` interval of 1000ms the newly-formed IPv6 address is considered to be unique.

For the purposes of IP-roaming state establishment, in particular sCoA generation, the IPv6 CoA must be generated off-link at some HAR neighbour (HAR_n), while the MN resides on the link of AR_c, as shown in Figure 4.8(b). Each HAR neighbour, generates *statelessly* the IPv6 sCoA on behalf of the MN, identified under proactive IPv6 handoff management as *proxy-stateless address auto-configuration.* In a manner similar to stateless address auto-configuration, the HAR neighbour AR_n, subjects the generated sCoA address to a DAD check: HAR_n first checks its neighbour cache against that sCoA to see if it is already existent; if not, it creates an entry with the sCoA and link layer address of the MN. The entry is marked with a P flag as *proactive* and set to `INCOMPLETE` state.

It can be seen that the only difference between proxy stateless address auto-configuration and the standard stateless address auto-configuration, is that the address is *proxy-autoconfigured* by AR_n. It does not require stored state or lease-lifetime configuration and hence, does *not* qualify as a stateless sCoA configuration through DHCP [112].

HAR_n then proceeds to complete DAD through standard address resolution: it sends a Neighbour Solicitation to the solicited-node multicast address mapping to the sCoA. It further includes as source Link Layer address option its own one on as the sender[6]. The sCoA address is found to be unique if no neighbour advertisement is received back within `RetransTime=1000ms` interval.

As soon as uniqueness of the soft CoA is ensured, the HAR neighbour marks the respective neighbour cache entry as *proactively reachable* or `P-REACHABLE`. This is a new state in Neighbour discovery [107] introduced for the purposes of enabling proactive reachability on that neighbour cache entry, representing the MN; in this manner, the particular entry *does not require* a solicited neighbour advertisement by the MN. This is because a neighbour solicitation is used in two broad occasions:

- when neighbour reachability or DAD is effected. The last degrades to address resolution which has the same effect on-link. Here the HAR neighbour is aware that no packet has arrived, to enforce a reachability test through a neighbour solicitation.

- when a packet arrives at the HAR neighbour for some host on its link, but its corresponding neighbour cache entry must be set to `REACHABLE`. The fundamental difference with the above case is that the HAR is aware that it acts in response to a packet that has arrived for the host on-link.

In both cases, the neighbour solicitation requires the link-local and L2 address of

[6]this function is usually performed by the MN when on-link

the MN. For each of the two cases the proposed model reacts differently. In the first case the HAR *defends* the sCoA with its own link-local and L2 address, by means of standard proxy-neighbour discovery; in the second case it simply returns the actual link-local and L2 address of the MN. However, the second case cannot occur unless some packet is routed towards the particular unicast sCoA; no CN or the HA knows the sCoA generated before the completion of MN's IPv6 handoff, typically revealed through a binding update.

By setting the link-local and L2 address of the MN in its neighbour cache while defending the entry with its own addressing, the HAR neighbour requires minimal information to activate the MN's link-local address; under IPv6 an active link-local address is typically required for forwarding to and from MN's particular soft CoA over the last hop.

Both MN and the HAR neighbour are configuring their neighbour cache entry in advance of IPv6 handoff with the particular entries set in the *P-REACHABLE* state. In this manner, communication between the two entities does not require a solicited neighbour advertisement; it can effect communication of packet traffic *immediately*. The P-REACHABLE state of the cached entry is reduced to REACHABLE as soon as the MN has sent a Binding Update (BU) to its peers. The BU ensures that a stable primary CoA has been activated.

E.6.2 Collection of IP-Roaming state at AR_c

With DAD completed, each HAR neighbour returns its own IP-Roaming state contribution though a unicast a *CtS-Avail* message, back to AR_c of the MN. Each state tuple received by AR_c, is grouped together in a sCoA tuple comprising the requested IP Roaming state of the MN.

Note that generation of soft CoAs is distributed among AR members of the R-neighbourhood, while the DAD function is performed at the HAR neighbour link during sCoA-generation time. Thus, no messages need to be *proxied* between AR_c and its AR neighbours for either CoA generation or DAD, as is the case with Fast MIPv6 handoffs. Such architectural decision saves a minimum of 1 RTT between the MN and the candidate HAR neighbour with the longest RTT.

The AR_c subsequently *injects* the state established, through a single *CtS-Response* message, onto the visiting MN for the type of context originally requested. In this manner the MN attains addressing and routing for the entire R-neighbourhood, ahead of its next IPv6 handoff as shown in figure E.6.

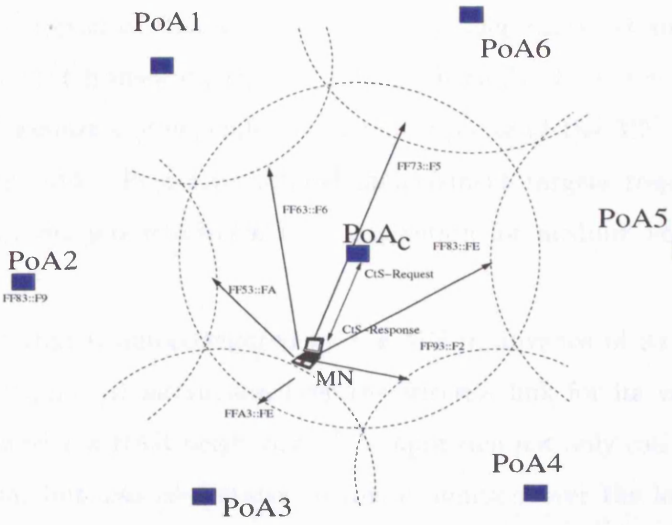


Figure E.6: The proactive set of soft CoAs provides 1-domain lookahead network connectivity

In addition to the sCoA tuple, the CtS-Response message may also include the following information per AR neighbour:

- the *prefix length* for that neighbour HAR interface. This is essential to derive link layer information for updating the MN's neighbour cache. It is required to minimise or eliminate neighbour discovery signalling when the MN handoffs to some AR neighbour.
- the AR neighbour *link layer address.* It is derived by using the prefix length together with standard EUI-64 rules for interface identifier generation; alternatively it may be explicitly provided by the AR_c as stored in its RNV Cache. It is used in the neighbour cache of the MN, marked with the P flag (PROACTIVE) and set in the `P-REACHABLE` state.
- the AR neighbour's *link-local address.* Such address is essential to the MN for the purposes of responding to neighbour solicitations initiated by the AR neighbour.

Hence the PoA tuple of Section 4.8 may be augmented to:

$$CtS_{AR_i} = ('Roaming', sCoA_i, DefaultRoute_i, PrefLen_i, MacAR_i, LLar_i, APID_i) \tag{E.9}$$

where $PrefLen_i$ is the length of network prefix identified by $DefaultRoute_i$ and $MacAR_i$ is the link-layer and $LLar_i$ the link-local address of the AR.

Despite the appeal of stateless address auto-configuration at the MN, proactive handoff management framework *refrains* from such mode of state configuration in an effort to *guard* against any dependence on the response of the MN back to the AR_c over the wireless link. Proactive handoff management targets responsive signalling interactions that remains *unaffected* from contention for medium access over the air interface.

An address that is autoconfigured by the MN in advance of its IPv6 handoff, is guaranteed to require (i) interaction over the wireless link for its validity (ii) proxy communications with a HAR neighbour. Such approach not only can be *influenced* by MAC contention, but also *contributes* to this contention over the local air interface. This is currently the modus operandi of Fast Handoffs for MIPv6. Clearly, in densely MN-populated coverage areas MAC contention effects become particular important as elaborated in Chapter 2 and shown by the work of Montavont and Noel [180].

In addition, as presented in chapter 6, MN-controlled proxy signalling introduces additional *overheads* for mobility-control interactions pertaining to address configuration, neighbour discovery, DAD or other context-specific action between the MN and the HAR neighbours

E.7 Statistical filtering of outliers from trace output

Before arriving at any statistical measure of central tendency (location) we identify any potential outliers from the produced trace output from individual simulation runs. This is significant, since outliers are bound to affect both confidence intervals and estimates of central tendency subsequently derived.

A typical method of determining outliers is to identify whether the output dataset of each simulation run is derived from the same population, by observing whether each run follow the same mean. To achieve this, dataset traces are first subject to a normality test. Figure E.7 plots the handoff delay values against a normal probability plot; the discontinuity in the plot shows clearly that datasets are not normally distributed. However, piece-wise normality appears to be hinted for the set of proactive handoffs (lower left) in contrast to the reactive handoffs (upper right)

Handoff delay traces are, thus, further split between reactive and proactive delay components to identify whether the individual handoff delay components are normally distributed. Piece-wise normality effectively implies that reactive handoffs are considered to be the transient outliers of the measurement as the handoff AR neighbourhood

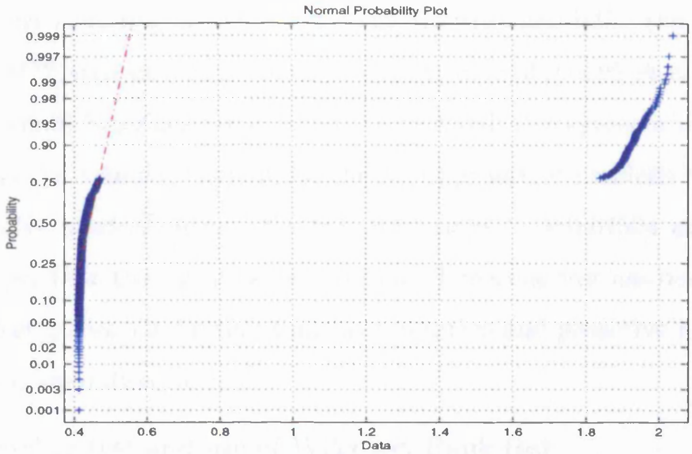
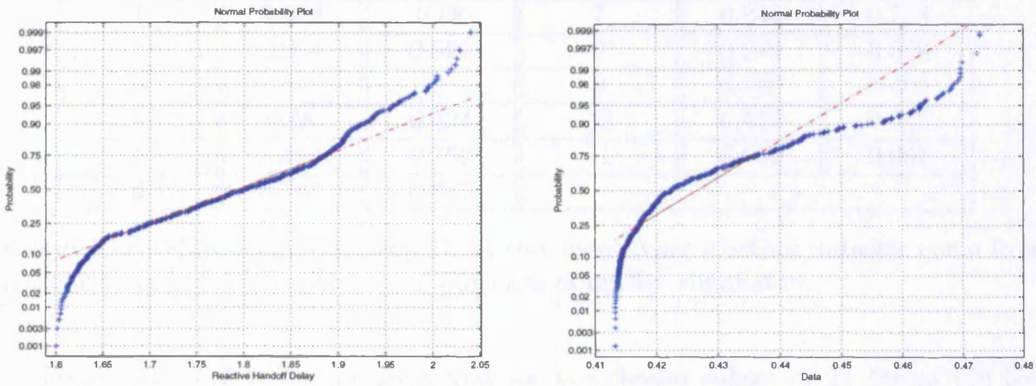


Figure E.7: Evaluation of handoff delay output trace versus a normal probability plot. The composite trace of handoff delay performance clearly is not normally distributed

gets discovered.

The normality test of both reactive and proactive handoff delay components is shown in figures E.8(a) and E.8(b). For the reactive trace data subset, it appears that samples at the centre of the dataset follow a fairly accurate linear pattern. This is not however the case for tails of the distribution. This indicates that the normal distribution is not the correct fit for the reactive handoff delay component. For the proactive handoff delay component there exist no linear relationship with a normal distribution; hence proactive handoff delay is clearly not normally distributed.



(a) reactive handoffs component of trace output

(b) proactive handoffs component of trace output

Figure E.8: Representative Probability plots for reactive handoff and proactive handoff samples. Irrespective of the removal of the handoff delay outliers (reactive handoffs) the set of proactive IPv6 handoffs is not normally distributed

To verify the above a Lilliefors test is further conducted assessing numerically the goodness of fit to a normal distribution. For reactive handoffs, the respective test statistic $l_s = 0.0779$ appears larger than the cutoff value of 0.0395 at a 5% significance level, hence the (null) hypothesis of normality is rejected; this agrees with the respective p-value = 0.0034. In a similar fashion, for the set of proactive handoffs before and after removal of reactive handoff outliers the test statistic is $l_s = 0.04524$ and $l_s = 0.04167$ which is still larger than the cutoff value of 0.0395. The same test has been performed in all remain dataset traces, confirming that both reactive and proactive handoffs dataset do not follow a normal distribution.

Failure of normality test and use of Wilcoxon Rank test

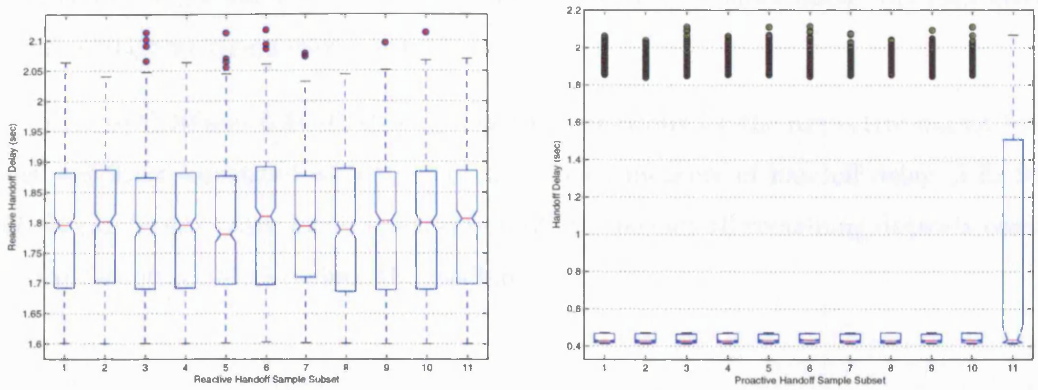
Since the aforementioned dataset is not normally distributed the outlier assessments necessitates the use of a Wilcoxon Rank test, to identify whether the two samples come from a common (non-normal) population. For p-values significantly greater than zero we accept the null hypothesis, and hence confirm that the two samples come from the same population. For p-values near zero we accept the alternate hypothesis implying that the data do not come from the same population, that is one of the two is significantly different and thus does not represent the measurement population expected. Such dataset is eliminated from subsequent statistical analysis. The p-value of the Wilcoxon Rank test are shown in table E.2.

Sample Id	Reactive	Proactive	Sample Id	Reactive	Proactive
	Wilcoxon p-value			Wilcoxon p-value	
1	0.652	0.88	7	0.844	0.788
2	0.254	0.607	8	0.979	0.588
3	1	1	9	0.789	0.404
4	0.48	0.424	10	0.479	0.452
5	0.426	0.769	11	0.652	0.061
6	0.784	0.677			

Table E.2: P-values from Wilcoxon Rank test identifying whether datasets come from the same sample population, for the purposes of outlier elimination

From table E.2 it can be seen that for the chosen subset of 11 traces the last one appears exhibits a much smaller probability (near zero) of belonging to the same population as the other samples. The results of the test statistic are visualised through the box-plots of figures E.9(a) to E.9(b).

The lower and upper lines of the 'box' are the 25th (Q1) and 75th (Q3) percentiles of the sample. The distance between the top and bottom of the box is the interquartile range (IQR = Q3-Q1). The interquartile range is a more stable measure of spread or



(a) Reactive Handoff

(b) Proactive Handoff

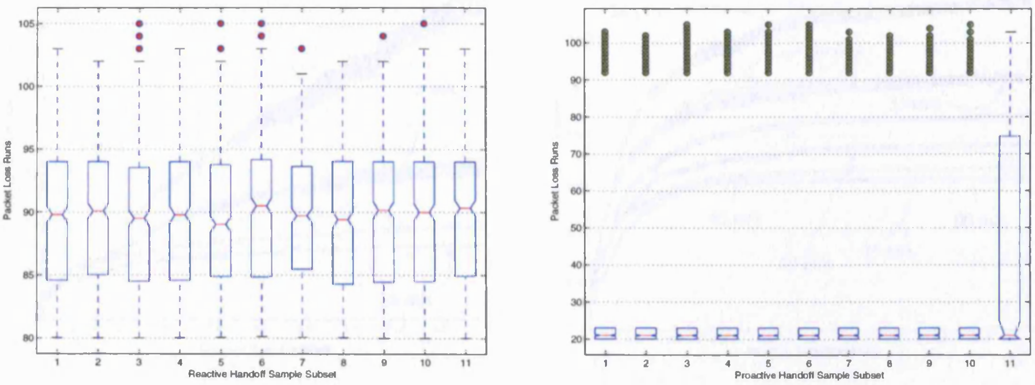
Figure E.9: Handoff Delay distribution of the upper, lower and inter-quartile range, including perceived outliers

dispersion [427, 428] in comparison to sample means. The line in the middle of the box indicates the sample *median*. If the median is not centred in the box (i.e. not identical to the mean), that is an indication of skewness. The 'whiskers' show the extent of the rest of the sample (unless there are outliers). In the case of no outliers, the maximum of the sample should be the top of the upper whisker. The minimum of the sample is the bottom of the lower whisker. By definition, an outlier is a value greater by about 1.5 times the interquartile range away from the top or bottom of the box. Points at the top of the plot indicate such data outliers. The 'waistline' median indicates graphically at its top and bottom the respective confidence intervals for each of the sample subset taken from the simulation measurement.

While in the case of the reactive handoff, the test statistic (its p-value) shows that the null hypothesis can be accepted (i.e. all generated samples are emerging from a homogeneous sample population (i.e. no simulation artifact or significant outliers), this is not the case for all trace samples of proactive MIPv6 handoff delay component. In particular, the $11-th$ dataset while it maintains a median nearly identical with the rest of the samples, it experiences significantly more reactive IPv6 handoffs during Handoff AR discovery.

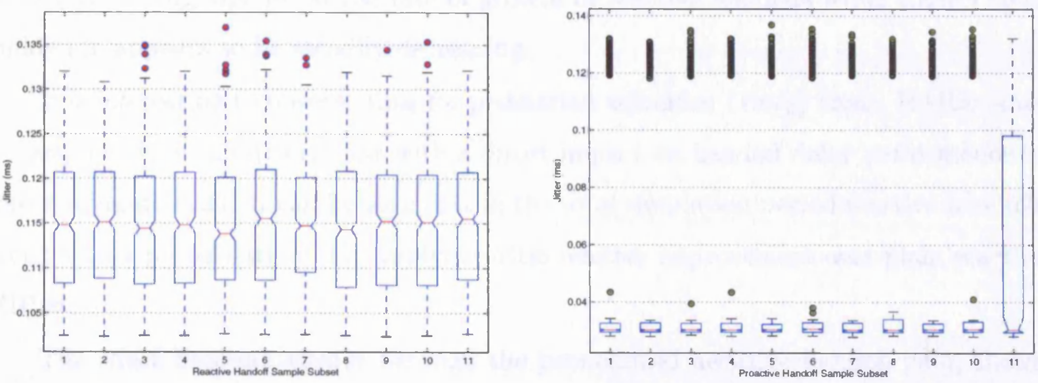
This behaviour has been investigated in the remaining 19 (out of 20) sample traces and it appears to be an artifact in the generation of the mobility pattern of MN. In particular, MN trajectories that pass with a less than uniform probability near the boundaries of the grid, appear to delay the discovery of Handoff ARs. This incurs a significantly larger amount of reactive handoffs, before the state about the respective Routing-Neighbourhood starts to build up.

Figures E.10 and E.11 illustrate percentile behaviour for the respective packet loss runs and jitter associated with the aforementioned measure of handoff delay of figure E.9. It can be seen that with the exception of one dataset all remaining datasets come from the same population since the median



(a) Reactive Handoff (b) Proactive Handoff

Figure E.10: Packet Loss Runs distribution of the upper, lower and inter quartile range, including perceived outliers



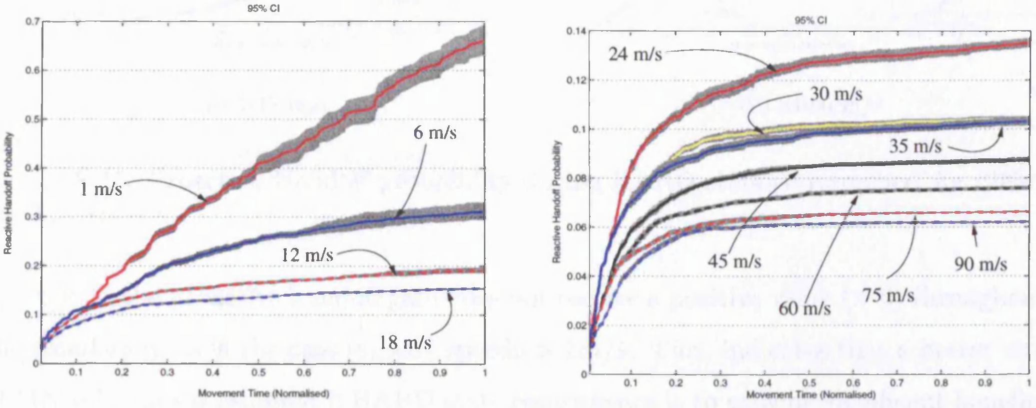
(a) Reactive Handoff (b) Proactive Handoff

Figure E.11: Jitter distribution of the upper, lower and inter quartile range, including perceived outliers

E.8 Parameters of influence in HARD convergence

E.8.1 Influence of Mobile Node Speed

To assess the effect of a varying measure of speed, the above simulation setup was augmented to an extended set of MN trajectories for the 10 participating MNs, with a maximum measure of MN speed varied between 1 and 90m/s[7]. All plotted curves, co-plot the respective 95% confidence interval.



(a) 1-18 m/s (b) 24-90 m/s

Figure E.12: Reactive Handoff probability during HARD state convergence, for different MN speeds (scales zoomed in where necessary)

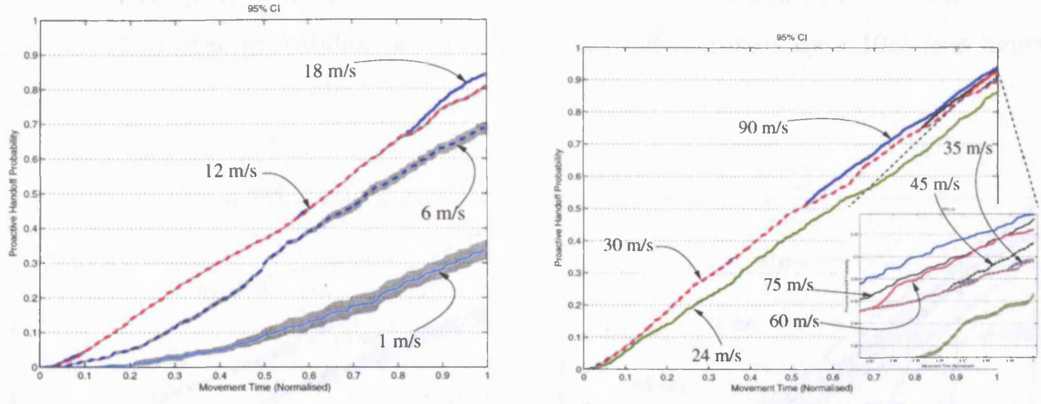
Figures E.12 and E.13 show the performance of reactive and proactive handoff ratio for a monotonically increasing maximum MN for each individual set of simulation runs[8]. For an increasing MN speed the rate of growth of reactive handoffs while HARD state builds up, appears to be steadily decreasing.

It is interesting to observe that for pedestrian velocities (1m/s) alone, HARD state convergence is prohibitively slow with a direct impact on handoff delay performance as shown in figure E.20. It can be seen that in the total simulation period reactive handoffs occur with a probability of 0.7, implying little relative improvement over plain reactive MIPv6.

The effect becomes clearer through the pronounced negative handoff gain, shown in figure E.14; the graph shows clearly that at 1m/s the curve of negative handoff gain (due to reactive handoffs) maintains a positive value throughout the simulation, instead of becoming negative (< 0) as is the case for higher MN speeds (e.g. 6m/s). Likewise in

[7]for very high speed motorways available in countries like Germany

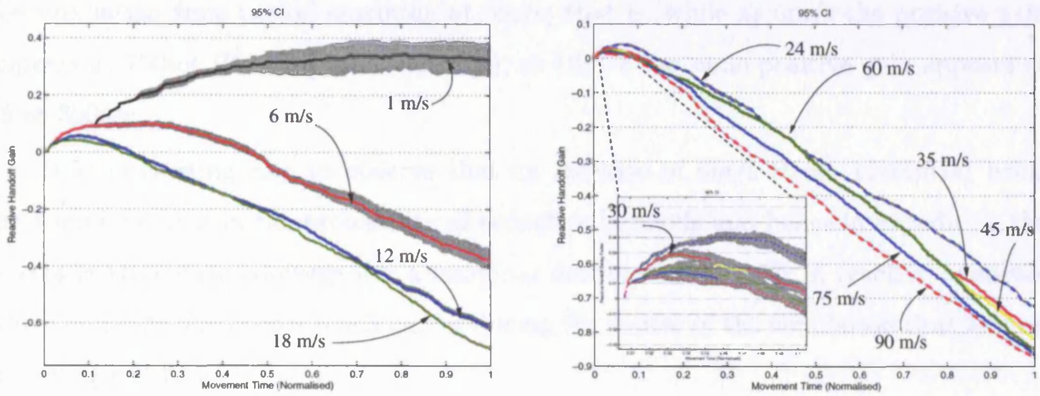
[8]20 iterations per speed value variation



(a) 1-18 m/s (b) 24-90 m/s

Figure E.13: Proactive Handoff probability during HARD state convergence, for different MN speeds

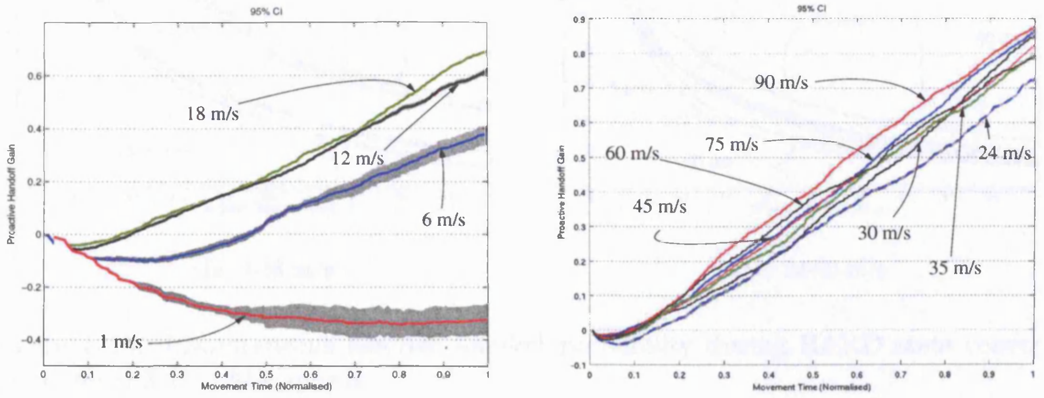
figure E.15 the proactive handoff gain does not receive a positive value (> 0) throughout the simulation, as is the case for MN speeds $> 1m/s$. This, indicates that a better mix of MN velocities is required if HARD state convergence is to provide significant benefits towards significant handoff delay reductions under proactive handoff management. In the context of these results, *better* implies higher MN speeds.



(a) 1-18 m/s (b) 24-90 m/s

Figure E.14: Reactive Handoff Gain during HARD state convergence, for different MN speeds

Indeed, performance as a result of faster HARD state convergence improves significantly for speeds $> 6m/s$. In particular about a 40% increase in proactive handoffs when the sparse set of 10 MNs maintain a speed mix with upper bound of 6m/s. From figure E.14 it may be seen that the rate of proactive handoffs appears to overtake its reactive counterpart at about half the simulation time. Doubling the maximum MN speed to 12m/s the probability of a proactive handoff increases by a 10% (see figure E.13).



(a) 1-18 m/s (b) 24-90 m/s

Figure E.15: Proactive Handoff Gain during HARD state convergence, for different MN speeds

However, the interesting improvement, as shown in figure E.14, is that the positive gain from the dominant proactive handoff delay component emerges at about 60% less the simulation time period emerging at 6m/s; that is, while at 6m/s the positive gain appears at 750sec (0.5 of simulation time), at 12m/s the same positive gain appears at 15 at 300sec.

It is interesting also to observe that an increase of 6m/s (from 12-18m/s) bring little improvement in the probability of proactive handoffs and hence in expediting the rate of HARD state convergence; a marginal decrease of only 5% in reactive handoffs, which none the less occurs much earlier during the course of the simulation time as seen from figure E.12.

An identical, albeit marginal, improvement is noted also for maximum MN speed between 24 and 30m/s, confirming that between 12 and 30m/s (i.e. a speed increase by a factor of 2.5) brings an improvement in HARD state convergence and subsequently in proactive handoff delay by about 10%. From there on proactive handoff gain gradually diminishes for speeds above 30m/s. In particular, an increase of 15m/s (from 30m/s) results an improvement in HARD state covergence of only 1.6%, where as a speed increase of 30m/s (from 60m/s) brings a reduction in reactive handoffs of only 1.5%.

This is a better-than-linear improvement in HARD state convergence, as illustrated

Max MN speed	Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
m/s [mph]-(type)	Handoff Delay (sec)					
1 [2.23]-(pedestrian)	0.468	1.920	1.934	1.940	1.956	1.959
6 [13.42]-(cycle)	0.424	0.465	1.910	1.924	1.942	1.956
12 [26.84]-(slow car)	0.422	0.437	1.881	1.919	1.941	1.966
18 [40.26]-(city)	0.422	0.433	0.466	1.903	1.940	1.972
24 [53.68]-((suburban)	0.421	0.433	0.462	1.898	1.942	1.962
30 [67.10]	0.421	0.431	0.444	1.877	1.925	1.951
35 [78.29]-(highway)	0.421	0.432	0.445	1.883	1.932	1.966
45 [100.66]	0.421	0.431	0.441	1.859	1.930	1.980
60 [134.21]	0.421	0.431	0.448	0.470	1.920	1.949
75 [167.77]	0.420	0.430	0.440	0.461	1.922	1.948
90 [201.32]-(autobahn)	0.421	0.430	0.440	0.456	1.915	1.942

Table E.4: Percentiles of handoff delay performance for varying MN movement speeds

of IP connectivity state established. For instance, for proactive establishment of IP-Roaming state the mechanism requires 1 RTT between MN's current AR and the farthest handoff AR neighbour; in addition, it requires 1000ms for DAD purposes; hence, the total cell residence period for IP Roaming state establishment purposes is bound between 1.08 and 1.16 sec for a typical RTT between 80-160ms.

Situations where the MN speed can outpace the transmission radius of the PoA over successive horizontal handoffs do not occur in practise since AP installation typically adapt their power budget and transmission range capabilities to accommodate practical maximum vehicular velocities attainable in the surrounding terrain [328].

Figure E.21 plots the respective probability density function of proactive handoff management through its respective proactive and reactive handoff probability densities for speeds of 1-30m/s (see figure E.22) and 30-90m/s (see figure E.23) respectively. It may be seen that for speeds of 1-30m/s the probability of a proactive handoff is 0.91; surprisingly even by increasing the maximum speed by a factor of of 3 (i.e. from 30 to 90m/s) the probability of a proactive handoff increases only by 5% to 0.96.

The implications of such performance behaviour is two-fold: (i) MN speeds above 24-30m/s do not increase significantly the rate of convergence of HARD state and subsequently the probability of a proactive handoff as a result of HARD state availability; (ii) MN speed alone is a significant factor in achieving fast HARD state convergence for an MN speed mix $\geq 6m/s$.

Figure E.24 shows the collective measure of proactive handoff probability as a function of MN speed expressing the rate of HARD state convergence for a sparse set of 10 MN moving within a topology of 45 PoAs and pause time 10sec. It may be seen that variance in the attained results decreases as the measure of speed increases.

Furthermore, the graph contrasts the performance of proactive handoff management in the light of varying MN speed measure versus the reactive MIPv6 standard; under reactive MIPv6 all handoffs are treated with reactive-type measures of delay (in the region of 1.6-1.9sec). On the contrary,under proactive handoff management the set of handoffs are mapped in terms of delay onto *transient reactive*, while HARD state builds up in PoA caches and proactive as supported by proactive IP Roaming (i.e. context-specific) state establishment.

It may be seen that for a maximum of 12m/s the probability of a reactive handoff drops from 0.7 down to 0.2; this is a 50% increase achieved for very slow vehicular speeds. At 30m/s (motorway speeds) the reactive handoff probability drops down to 0.1, with proactive handoff probability reaching 0.9. The above imply that for small maximum MN velocity measures, MN speed should be complemented by other factors (such as smaller pause periods or higher MN densities) if HARD state is to coverge significantly faster.

In the following section simulations investigate the effect of pause period of MN movement on the measure of proactive handoff probability.

manifested as an oscillating delay pattern, during HARD state convergence, between reactive and proactive handoffs. As the density of the topology increases (see Table E.7) convergence of HARD state appears to be more scattered throughout the simulation period. On the contrary, for a low PoA density (e.g. PoA=45) HARD signalling tends to appear early during simulation, compressing state convergence at the start of the simulation period. The density of the plotted histogram for PoA=1000 hints further about the significantly increased measure of handoffs performed in comparison to low PoA densities.

No of PoA	PoA Density	mean	median	trimean	Std.Dev	Min	Max
#	$\lvert PoA\rvert/m^2$	Handoff Delay (sec)					
45	1.1	0.654	0.430	0.484	0.529	0.415	1.971
70	1.7	0.711	0.433	0.559	0.577	0.415	2.009
140	3.38	0.743	0.436	0.602	0.599	0.415	1.950
200	4.8	0.804	0.436	0.681	0.638	0.415	1.978
300	7.25	0.775	0.436	0.644	0.622	0.416	1.995
400	9.7	0.811	0.435	0.689	0.643	0.416	1.965
500	12.1	0.829	0.437	0.715	0.652	0.416	1.977
600	14.5	0.816	0.437	0.697	0.643	0.414	1.9524
700	16.9	0.795	0.442	0.667	0.632	0.415	1.983
1000	24.2	0.821	0.437	0.704	0.647	0.415	1.980

Table E.7: Measures of location (central tendency) in handoff delay performance for varying PoA densities between 45-1000 PoAs serving a sparse set of 10 MNs

No of PoA	PoA Density	Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
#	$\lvert PoA\rvert/m^2$	Handoff Delay (sec)					
45	1.1	0.423	0.446	1.888	1.916	1.957	1.971
70	1.7	0.424	0.462	1.906	1.927	1.955	2.002
140	3.38	0.425	0.466	1.904	1.919	1.944	1.95
200	4.8	0.425	0.467	1.91	1.922	1.955	1.978
300	7.25	0.425	1.847	1.914	1.928	1.948	1.995
400	9.7	0.425	1.856	1.912	1.925	1.950	1.965
500	12.1	0.426	1.864	1.908	1.925	1.95	1.976
600	14.5	0.426	1.859	1.907	1.923	1.947	1.952
700	16.9	0.428	1.842	1.915	1.932	1.957	1.983
1000	24.2	0.426	1.856	1.909	1.929	1.958	1.979

Table E.8: Percentiles of handoff delay performance for varying PoA densities between 45-1000 PoAs serving a sparse set of 10 MNs

Tables E.7 and E.8 show the measure of central tendency and respective percentiles of handoff delay performance for varying PoA densities. From table E.8 that at 75-th percentile of handoff delay measure, ; it, thus, becomes clearer that for topologies above 200 PoAs at least 25% (but $< 50\%$) of the handoffs are expected to experience reactive-

It appears that the absolute minimum period of cell residence for an MN is the RTT between it and the farthest HAR neighbour allocating proactively state of any particular context. This implies that for an average worst case RTT of 160ms, the MN *cannot* transit between successive PoA with a handoff rate greater than 6.25 h/sec. This results are also in agreement with preliminary performance results in the study of Chalmers et al [147]; however, this study did not provide any results towards the tradeoff of cell residence period and the delay incurred by proactive state establishment before the next successive handoff for an MN.

From the above it is concluded that for a sparse set of (10) MNs operating in very dense PoA topologies, a handoff rate $> 1\ h/sec$ prevents proactive handoff management from supporting handoff delay seamlessness when the minimum cell residence period is $> 1\ sec$; this is the case when *proxy-stateless* address auto-configuration is employed for the purposes of IP Roaming establishment. On the contrary, statefull address autoconfiguration, enables significantly better performance for proactive handoff management since the measure of cell residence period can be significantly reduced when IP Roaming state context is established with HAR neighbours and thu,s enable conservative handoff rates of up to 6 $handoffs/sec$.

The reason for degraded proactive handoff performance is the slow HARD state convergence as a result of (i) sparse MN population and (ii) a significantly larger number of HAR neighbours per routing neighbourhood. Notwithstanding, the encouraging observation is that in a typical mobile environment typical MN populations within a single geographical area range between 10,000-50,000 subscribers. In addition, it is very rare[14] statistically, that all MN may communicate simultaneously. The following section in this study evaluates the effect of large numbers of MNs operating within a PoA topology.

E.8.4 Influence of the number of MN

To assess the effect of MN population size on the rate of HARD state convergence, we adapt the simulation model, to encompass an increasing number of MNs within the PoA topology. To this end, seven distinct simulation scenarios are identified with MN population size between 25-1000. The remaining parameters of importance, namely speed, pause period and PoA size, employ default nominal values: these are 10m/s, 10sec and 45 PoAs respectively. A number of 20 iterations is produced for each MN population

[14] the frequency of occurrence of such events is perhaps 2-3 per year including significant national holidays or states of emergency

requires an MN velocity of $\geq$ 60m/s. The former implies, that the MN population size emerges as the most influential parameter for HARD state convergence, while MN speed appears to be the second most influential parameter in the performance of proactive handoff management.

While the measure of reactive handoff probability (and, thus, its respective measure of delay) appears to remain less than 6% for high MN densities (e.g. 1000 MNs), figure E.41(a) illustrates a rather counter-intuitive effect: as the number of MNs increases from 25 to 200 the measure of reactive handoff probability appears to decrease monotonically with an (inversely proportional) linear rate of decrease; that is, as the number of MN doubles (from 25-50 and 50-100) the rate of decrease in reactive handoff probability halves (from 3.2% to 1.7%).

However as the number of MNs becomes > 400 (e.g. 500) the measure of reactive handoff probability becomes proportionally (monotonically) *increasing*, in a non-linear manner; for an increase of 200 MNs the probability increases by 0.3%, for a subsequent increase of 250 MNs, reactive handoff probability notes an additional increase of 1% while for the next (last) increase of 350 MNs it notes an additional increase of 0.3%. The effect magnifies as the number of MN increases asymptotically, but terminates within shorter initial transient periods during simulation.

The cause of such behaviour is sought in the sheer amount of MNs transiting simultaneously onto a new PoA, during their first few reactive handoffs. During these first simulation moments, a large number of MNs transiting between PoAs experience simultaneously, their first reactive handoffs. This causes the instantaneous ratio of reactive handoffs (over the total (running) measure of handoffs performed) to shoot at significantly high values of about 5% (for 1000 MNs - see figure E.41(b)) within a small period (0.01) of the simulation time (about 15 sec).

Such high jump of the total reactive handoff probability is only temporal (in effect for asymptotically small periods of time) at the start of the HARD state convergence period. Nonetheless, such effect inflates the measure of reactive handoff probability in its overall value, for an increasing number of MNs, as shown in figure E.41(b). Such measure, however, does not reflect the average value of reactive handoff probability throughout the simulation. The latter is reflected clearly in figure E.43(b), plotting the measure of reactive handoff gain and the measure of instantaneous reactive handoff decay of figure E.45.

Similar observations apply for the measure of proactive handoff probability shown

Number of MN	MN/m^2	mean	median	trimean	Std.Dev	Min	Max
#		Handoff Delay (sec)					
25	1/8100	0.492	0.424	0.425	0.305	0.415	1.977
50	1/4050	0.472	0.424	0.425	0.258	0.414	1.994
100	1/2025	0.461	0.424	0.425	0.226	0.414	1.996
200	1/1012.5	0.453	0.423	0.424	0.202	0.414	1.995
400	1/506.2	0.457	0.424	0.425	0.214	0.414	1.997
650	1/311.5	0.471	0.429	0.425	0.255	0.414	2.0033
1000	1/202.5	0.488	0.425	0.425	0.297	0.414	1.996

Table E.9: Measures of location (central tendency) in handoff delay performance for varying MN densities between 25-1000 mobile nodes over a topology of 45 PoAs

Number of MN	MN/m^2	Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
#		Handoff Delay (sec)					
25	1/8100	0.420	0.430	0.440	0.463	1.931	1.972
50	1/4050	0.421	0.430	0.438	0.447	1.925	1.961
100	1/2025	0.420	0.429	0.436	0.444	1.914	1.954
200	1/1012.5	0.420	0.428	0.435	0.441	1.908	1.950
400	1/506.2	0.420	0.429	0.436	0.442	1.916	1.962
650	1/311.5	0.420	0.430	0.438	0.445	1.925	1.966
1000	1/202.5	0.421	0.430	0.441	0.461	1.933	1.976

Table E.10: Percentiles of handoff delay performance for varying MN densities between 25-1000 mobile nodes over a topology of 45 PoAs between the communicating peers. Conversely, small amounts of fixed play-out delay improves interactivity, but incurs higher packet loss in the play-out buffer, degrading the quality of VoIP speech output. The value of T_{e2e} must be chosen based on some knowledge of the delay in the network. However, such metrics may not always be available and/or the statistics of the network delay may vary with time. To this end, adaptive play-out schemes are considered.

Adaptive play-out schemes monitor the network delay and its variations and adjust accordingly the play-out time of voice packets. A number of algorithms have been proposed managing delay on a talkspurt-by-talkspurt basis; [429] investigates the monitoring of network delay estimating both delay and its variance, using moving averages, to adapt the play-out time at the beginning of each talkspurt but keeping it constant throughout the talkspurt. The scheme improves by detecting delay spikes, adapting faster play-out delay during the spike periods. The latter has been adopted by a number of subsequent schemes proposed in [430], exhibiting improved performance by employing percentiles instead of moving averages, at the cost of increased state and processing.

A different group of play-out algorithms adapt play-out delay on a per-packet as opposed to a talkspurt-by-talkspurt basis, allowing for tracking of delay variations even within a single talkspurt; [294] adopts such a technique, but does not take into account the voice signal itself, with an impact on the pitch of the speech signal during play-out. The work in [298] adjusts such effects by scaling voice packets through a time-scale modification scheme.

It may also be noted that, most play-out algorithms incur or allow for some packet loss. For instance, it is observed that managing delay through moving averages, causes packets with delay value located at the tail of the distribution to be lost. To this end, most of the schemes allow also to specify a target loss rate [298, 431, 432, 430, 298]

Given the interactive nature of voice conversations, it is intuitive that voice packets arriving after the scheduled play-out deadline are too late to support continuity in the uttered speech signal at the receiver and thus must be discarded. That is to say, *delay in the receipt of packets translates to late loss rates* over interactive real-time IP services.

F.2 Empirical measure of wireless handoff rate

For robust protocol design in IPv6 mobility management, it is important to identify the average number of handoffs, occurring during a voice call effected of wireless networks.

To this end, we utilise data drawn from the realm of cellular circuit-switched communications. Such information, provide a realistic measure, indicative of how frequently an IPv6 handoff is expected to occur (assuming a different subnet per PoA) and thus, disrupt the flow of communication of the MN with its peers while in transit.

Holub et al [433] provide a short statistical analysis of average call duration, derived from the quality call record database of 1,000,000 calls of an operational GSM network. The authors report a median (50-percentile) call duration of 107 sec. Their derived empirical probability distribution function of call duration appears to resemble a gamma distribution whereby, 95% of the calls complete within 350 sec, while only 3% and 2% of the calls complete within 400 and 450 sec respectively. Nanda [273], in his teletraffic model analysis employs an average call duration of 150 sec derived from empirical data, while Gavish and Sridhar [434] converge to the median call duration identified by Holub, using an average call duration of 120 sec based also on empirical data. The above indicate an average call duration with range between 120-150 sec.

With respect to handoff frequency during a call, Feuerstein et al [435] shows experimentally that the measure of average number of handoffs depends on the measure of hysteresis of the wireless terminal. The authors show that for intra-city cell ranges of about 275m, the mobile terminals register over a cellular TDMA system, on average 5 and 20 handoffs during a *500-sec* voice call, for typical hysteresis levels of 7 and 3 dB, respectively, at pedestrian speeds (0.8-1.2m/s). Approximating these values onto the average call duration identified above, indicates an average frequency of 1-2 handoffs for a high and 5-6 handoffs for a lower hysteresis margin, per call.

Nanda [273] provides a different classification of handoff rates over production cellular networks, by looking at the relation between cell size and the respective measure of traffic load.

Their results are based on the observation, that in the face of increased traffic load, WISPs provide more capacity by reducing cell size, while populating the same geographical area with more base stations [436]. The increase in system capacity is proportional to $1/r^2$ where r is the cell radius; for instance, a reduction in cell radius by a factor of 4 (e.g. 4 to 1 km) would result in a 16-fold (i.e. 4^2) increase in capacity.

Their results suggest that an increase in traffic load causes a reduction in the effective cell size, through the deployment of shorter range[1] base stations that cover the same geographical area. Such management decision has a knock-on effect on the

[1]and thus non-interfering

number of handoffs per voice call.

In particular, for intra-city cell sizes between 395-500m and a congestion factor $c = 0.1$, Nanda reports that a pedestrian MN experiences approximately 2-3 handoffs/call. As the congestion factor increases to $c = 0.5$ the cell size reduces from 395 to 158m with the handoff rate for fast moving vehicles (25m/s) of over 20 h/call. For fast-moving vehicles over macro-cells the mean handoff rate reaches about 4 handoffs per call. Interestingly enough, as the number of slow (pedestrian) MN increases ($0.4 \geq c \geq 0.8$), the average number of handoffs experienced by slow MNs increases to about 3-5 handoffs per voice call. Thus, for an average congestion factor of $c = 0.5$, fast moving MNs observe an average handoff rate of 8 $h/call$, while slow moving MNs experience an average handoff rate of 4 $h/call$. Table F.1 summarises the operational measure of handoffs/call as reported in results of [273].

cell size	MN speed	congestion factor	handoff rate
m	m/s	no of MN/cell	handoffs/call
395-790	1m/s	c = 0.1	2-3
158-395	25m/s	c = 0.5	> 20
790	25m/s	c = 0.5	7-9
1000	25m/s	c = 0.5	4
1000	1m/s	c = 0.5-0.8	3-5

Table F.1: Empirical handoff rates observed in operational Cellular networks

The above imply that for pedestrian speeds (1m/s), a voice call would experience at the MN/CN on average, a flow disruption (as long as the measure of persistent handoff delay), every 30-37.5 sec for a call duration between 120-150 sec. For vehicular speeds (up to 25m/s) a voice call would experience at the MN/CN a flow disruption every 15-18.75 sec. The measure of disruption becomes more critical for fast moving MNs, since packet loss detracts more frequently from the intelligibility of the voice conversation. Thus, for both MN-speed classes, the measure of persistent handoff delay establishes a noticeable degradation in voice communication at a significant frequency.

For IEEE 802.11 WLAN networks, the measure of handoff rate becomes more pronounced, given that a typical maximum cell size $< 300m$. Henderson et al [306] report on the monitored usage of an intra-campus WLAN network, that wireless nodes appear to associate and re-associate with several APs many times in succession, despite the host's physical *stationarity* within the AP cell.

If 802.11 MNs perform such measure of L2-handoffs while stationary, it becomes obvious that the average handoff rate per voice call would increase significantly. Indeed the authors report an average AP visit per MN trip of 12 APs. Had the WLAN environment supported IPv6 mobility management and assuming that each AP is attached to an individual AR, each MN would experience an average IPv6 handoff rate of 12 h/call, with a reported call duration of 31 sec. The former implies a potential VoIP flow disruption every 2.58 sec, with a persistent handoff delay in excess of 350ms per handoff.

From the above observations it emerges further that IP flow disruption effects are exacerbated, when the MN considers apart from horizontal, also *vertical* handoffs between competing WISP wireless cell overlays, as a result of higher handoff rates. A vertical handoff may occur as a result of multi-PoA diversity while the MN is stationary, as opposed to physical MN movement beyond the bounds of a coverage area.

F.3 L2-handoff delay in IEEE802.11b/g WLANs

The process of a link-layer handoff has been elaborated in detail in Chapter 3; experimental measurements over a particular WLAN vendor implementation exhibited a significant measure of L2-handoff delay, well over 200ms. This section provides a passing view of the state-of-art in schemes or recommendations aiming to reduce L2-handoff latency. We focus on the trade-offs that such mechanisms make establishing the rationale for proposing subsequently particular generic L2-handoff optimisations extending the performance of proactive handoff management.

Before looking at trade-offs of existing solutions, it is important to note that the IEEE 802.11 specification [80, 17] prescribes only the *mechanisms* to implement the link-layer handoff. The duration of the L2-handoff is not specified, allowing implementation vendors to balance trade-offs between fast response and low power consumption.

It is further important to highlight the process of L2-handoff initiation: during a frame transmission failure, the MN assumes at first stage, frame *collisions* and retransmits several times using lower bit rates (through dynamic rate shifting). If transmission remains unsuccessful, then *radio fading* is assumed and a probe request is sent to check the link. After *several* unanswered requests, the station declares itself in out-of-range status and thus, commences an *active scan.*

During an active scan each channel is probed briefly for active APs. The dwell period for each channel scan is bound by [`MinChannelTime`,`MaxChannelTime`]. If the channel is *idle* for MinChannelTime, namely there is neither response nor any kind of traffic in the channel, the scan in that channel is complete and the channel is declared empty. If there is any traffic during this time, the station must wait `MaxChannelTime`.

`MaxChannelTime` should be large enough as to allow the AP to compete for the medium and send the probe response. It is noted that a scanning station is not able to sense other stations communicating with the AP, but can receive the acknowledgements sent from the AP. Hence, unless the channel is idle the station always waits for `MaxChannelTime` for probe responses. While this period may vary per vendor implementation, a `MaxChannelTime` of 38ms per channel may be assumed, as identified in Section D.6.1 and confirmed by [303].

Yokota et al [105] propose a method for expediting an 802.11 L2-handoff that employs an inter-AP link-layer protocol and a dedicated MAC bridge. The protocol is based on broadcasting of MAC addresses from candidate APs over local WLAN segments connected through the MAC bridge. An AP broadcasts its MAC address when the MN performs an association (during an L2-handoff) with it. The MAC bridge is responsible for redirecting a flow of frames from the old to the new AP, achieving thus an L2-handoff. This proposal while appealing at local segments, appears to suffer from significant limitations. First, the solution assumes that all neighbouring APs can be connected through a single MAC bridge so that the MAC address broadcast by the AP can be feasible. Normally, this is not practical, since the WLAN deployment suggest multiple competing WLAN domains over a single geographical region. A number of AP neighbours would be connected over MAC bridges that span across *different* IP networks. Hence, for spanning tree last-hop WLAN networks a broadcast would not work. Furthermore, it is unlikely that WISP would allow redirection of packets at the MAC layer between domains, as this would require additional: (i) configuration effort, (ii) L2 devices (iii) topology-specific assumptions, (iv) security considerations.

Misra et al [409] propose a proactive caching of neighbouring APs through an inter-AP protocol. The scheme is based on WLAN re-associations being enriched with the identity of the previously visited AP, passed by the MN during its next AP association. This is essentially similar to proactive handoff management presented in Chapter 3, but applied exclusively to the link-layer. This mechanism appears to offer significant reduction in L2-handoff delay. However, it presents a number of inconsistencies on claimed performance that casts shadows on the accuracy of the reported results.

The authors report a re-association latency reduction from 15.37 to 1.69ms. From an implementation perspective, an L2-re-association latency of 1.69ms is impossible for the simple reason that link-local (ping) latency has been experienced to be 3.8ms[2] during measurements (see Section D.6.1, Chapter 3). In addition, as shown in Table D.2 of Section D.6.1, authentication incurs by itself a delay of 3-5 ms. Ultimately, the authors provide no information about the configured dwell time per channel scan (i.e. `MaxChannelTime`), to identify the duration of a single channel scan.

What's more, for APs with more than one neighbour, the MN would have to scan at least two channels. With very optimistic channel dwell periods of 5ms this implies a minimum of 10ms before the MN can proceed to the authentication stage. Furthermore, the lack of signalling sequence and the delay cost of each signal during the active scan in this mechanism makes impossible to confirm the claimed L2-handoff latencies.

Ultimately, the reported L2-handoff delay figures encompass a maximum of 3 operating channels. From monitoring measurements conducted over real WLAN network infrastructure[3], we found that nearly all 13 frequencies were used in the deployed Distributed Service Sets. The former implies a significantly higher L2-handoff latency. We are currently unable to assess how long this latency is since the maximum dwell period is not provided by the authors.

Velayos and Karlsson [437] propose a modification to the 802.11 L2-handoff mechanism whereby the probe process commences as soon as collisions have been excluded as a reason of frame transmission failure. The process meets significant complexities, however, when the wireless station is *not* transmitting. Furthermore, the authors propose that the active scanning should not scan all available channels, but do not suggest how to derive a shorter list of channels, or under what criteria should such channels be selected over others.

Ultimately the proposal by Sharma et al [438] encompasses interleaving during transmissions between infrastructure and active scan mode for the purposes of detecting APs while the MN is engaged in communication with its peers. The authors claim an L2-handoff delay of 60ms under such scheme with a probing interval of 10ms. In addition they report a delay of 100 μs per invocation.

These figures are, however, conflicting with results reported in similar implementation efforts. In particular, Chandra et al, in a similar implementation called MultiNet [336], report that each channel switching *requires* from the wireless interface a delay of 25-30ms. Ramani et al report under the Syncscan implementation [335], that it is

[2]at 1 Mbps and 2.14ms at 11 Mbps
[3]Intra-UCL Computer Science campus neighbouring with the UCL-Hospital WAN network

unrealistic to effect a scan more frequently than every 500ms since jitter and packet loss grow prohibitively for interactive real-time applications.

While interleaved scanning emerges as a significant proposal towards reduction of the search delay in WLAN L2-handoffs, its performance appears to be also hindered by practical limitations in terms of switching delays between scanning and communication as well as buffer exhaustion on the AP, or attendant packet loss [354].

F.4 Optimising 802.11 Scan latency

Given that L2-handoff performance must support bounded delay performance in this section we devise a simple scan latency optimisation for the L2-handoff sequence of 802.11. Such optimisation is assisted by the Handoff AR discovery (HARD) function of proactive mobility management.

In Section 4.7.1 we specified that a PoA member within the M-neighbourhood is identified by means of an MNV-element (Eq. 4.2, defined as:

$$MNVe_{k,i} = (APID_i, Channel_{APIDi}, \theta_{APIDi}, DomainID_l) \tag{F.1}$$

For the purposes of an L2-handoff the above contains two vital information components: (i) $APID_i$ which is the identity of the AP and (ii) $Channel_{APIDi}$ the operating channel of the AP.

We recall from the findings of Section D.6.1, that during an 802.11 L2-handoff, the AP discovery phase, produces the largest contribution to the total measure of L2-handoff delay. During this phase, the MN scans actively all available channels, with a minimum dwell period spend in probing each channel for operating APs. If the dwell period can be reduced to the absolute minimum by *guiding* the active scan process, then the dominant delay component of a WLAN L2-handoff could be reduced significantly with no impact to the robustness of the scanning process. If such reduction can ensure that the total delay incurred by an 802.11 L2-handoff remains well below 100ms, then the HandoffCast function can address realistically delay seamlessness during MN's IP handoff over WLAN.

To this end, the state established during HAR discovery at MN's current PoA, augments MN's proactively established IP Roaming state with *both* the list of AP identifiers *and* their respective list of operating channels. On receipt, the mobility management function of the MN inserts the list for the following pair through a reverse L2-hint back to link-layer of the wireless NIC. In this manner, the link layer is informed *explicitly* about which channels to scan.

Looking again at the figures of Table D.2 or Table D.3 of Section D.6.1, we may observe that irrespective of the size of Probe Request Interval, any number of APs residing onto a single 802.11 'channel' produces an average Probe response delay of 3ms at 1Mbps and 4.5ms at 11Mbps. This implies that under average MN association load any number of APs operating over a single channel would effect their probe responses within 4.5ms. If all 13 channels[4] are scanned with a maximum dwell time of 5ms, all APs neighbours would provide their probe responses within 65ms to the MN.

Together with authentication and association signals this would bring the total measure of L2-handoff delay to 107ms for the worse case scenario, where all of 13 channels need to be scanned[5]. With guide the results of Section D.6.1, we can anticipate through such cross-layer optimisation a significant reduction in L2-handoff delay. The latter clearly emerges as a potentially viable solution in 802.11 networks towards support of delay seamlessness through flow forwarding. We evaluate such expectation through simulations in combination with HandoffCast efficiency and performance.

F.5 Assessing optimised 802.11 L2-handoff performance

To assess the performance of the proposed L2-handoff optimisation, we perform a set of 802.11 MAC layer simulations through NS-2. A set of 10,20 and 40 MNs are placed randomly within a PoA topology similar to the one of Section 4.9.3. Wireless stations transmit a constant bit rate (CBR) flow with a packet payload of 33 bytes. The beacon period for all APs is assumed to be the nominal measure of 100ms.

Parameter	Value
Slot_Time	20 usec
SIFS	10 usec
DIFS	20 usec
CWmin	31
CWmax	1031
MinChannelTime	1.52ms
MaxChannelTime	30ms
Tx Rate	1 Mbps
MPDU Payload	33 bytes
Beacon Interval	100ms

Table F.2: Simulation Parameters for 802.11 L2-handoff optimisation

Before engaging into simulating the performance of the proposed optimisation, it

[4]ETSI allocation of channels in Europe. In North America a total of 11 channels would yield a complete scan within 55ms.

[5]In North America the total L2-handoff delay would reach 97ms

is important to attain a realistic measure on the bounds of `maxChannelTime`. This is because under loading conditions over a WLAN network, transmission of a probe request/response during an active scan is subject to transmission delays as a result of increased contention. The 802.11 specification prescribes a `maxChannelTime` of 30ms; however, this period may vary depending on the vendor implementation up to 38ms (see Section D.6.1).

From a mobility management perspective, the emerging question is what is the minimum measure of `maxChannelTime` that can accommodate an increased measure of contention when (i) the offered load increases (ii) the number of associated MNs increases.

To field such question we first look into the collision avoidance mechanism of 802.11, since it is this function that balances (frame) transmission delay with medium access and transmission robustness.

The collision avoidance mechanism at the MAC layer of the IEEE 802.11 specification [17] prescribes a binary exponential backoff period with a maximum backoff contention window of $CWmax = 1031$. Beyond this contention window, the wireless station must transmit its frame even at the cost of collision; this is to prevent starving the station from transmitting indefinitely due to high communication load.

Based on the contention-based backoff function, the minimum expected dwell time over an 802.11 transmission channel for the purposes of a single frame transmission is:

$$E[ChannelTime] = DIFS + (SlotTime * CWmin) + T_p \tag{F.2}$$

where T_p is the transmission period of the frame, $DIFS$ is a fixed DCF Inter-frame space period, and $SlotTime$ is the unit of the transmission period. The transmission period T_p is equal to

$$T_p = DIFS + PHYoverhead + T_d + ACK \tag{F.3}$$

where T_d is the transmission time of the data frame. For the size of a probe frame of 55 bytes, the T_p is found to be 900 μs. From the above we compute the expected measure of `maxChannelTime` as a function of backoff attempts.

Figure F.1(a) shows the number of estimated backoff attempts before $CWmax$ is consumed and the MN is forced to transmit its communicated frame.

It may be seen that the MN is effectively allowed 6 backoff attempts (counter resets)

$$t = \frac{2(1-2p)}{(1-2p)(W+1) + pW(1-(2p)^m)} \tag{F.6}$$

where m is the number of backoff attempts and W is the backoff window. This system has been validated and used in a number of 802.11 MAC throughput studies [441, 442, 443, 444].

Combined with eqn.(F.4) the non-linear system, provides the measure of access probability for an MN, as a function of the MN population within each AP cell; use of this model during simulations, allows to influence statistically the number of backoff attempts experienced by an MN, as a result of an increasing number of 802.11 stations.

Computing the probability of a successful transmission after k collisions as a system of equations F.4, F.5 and F.6, provides the maximum number of collision after the expiry of $CWmax$. Figure F.1(b) shows the incurred number of collisions before transmitting successfully a frame over a load-saturated 802.11 link. It may be seen that the number of collision does not exceed four, even for an increasing number of MNs, saturating the link capacity.

The above medium access probabilities are thus integrated in the devised simulation model, to account for the number of backoff attempts and respective number of frame collisions as a result of saturated transmission load. To attain a confidence interval the simulation scenario received 20 iterations. The MN employed the modified random way-point model of Section 4.9.4, moving within the nominal PoA topology of 45 APs. The derived measure of probe Request/Response delay is identified as the respective `maxChannelTime` for each channel. All PoAs are assigned randomly a channel number between 1 and 13. Measurements are acquired after HARD convergence period, which for a density of 45 PoAs was found to be 300sec. Hence, for a total simulation period of 1500sec the steady state period during which measurements were recorded, had a duration of 1200sec.

F.5.1 Simulation Results

Figure F.2 shows the measure of delay incurred by a probe request/response handshake during MN movement. It may be seen that for an MN population of 20 STAs and an average traffic load (50%), a probe response is received within 10ms (8.2ms). Hence, for the average load case of 20 STAs a maxChannelTime period of 10ms is sufficient for the detection of a channel. For a 70% load the maxChannelTime nearly doubles to 19ms. Above that load threshold $G > 0.7$, the probe response delay increases above 30ms.

signalling cost (Common,Exclude,Include) neighbours and PoA node-degree, observed during HandoffCast forwarding, as function of PoA density.

PoA density (ND)	mean	median	trimean	Std.Dev	Min	Max
#	Persistent Handoff Delay (sec)					
45 (8)	0.079	0.078	0.079	0.009	0.059	0.114
70 (6)	0.073	0.073	0.074	0.008	0.049	0.097
140 (2)	0.158	0.169	0.167	0.032	0.036	0.187
200 (3)	0.142	0.148	0.148	0.030	0.036	0.202
	L2-handoff Delay Component (sec)					
45 (8)	0.067	0.063	0.58	0.007	0.040	0.075
70 (6)	0.069	0.063	0.061	0.007	0.030	0.087
140 (2)	0.062	0.063	0.062	0.009	0.028	0.095
200 (3)	0.054	0.062	0.063	0.011	0.031	0.112
	Forwarding Delay Component (old→new PoA) (sec)					
45 (8)	0.040	0.040	0.040	0.008	0.011	0.067
70 (6)	0.032	0.032	0.032	0.008	0.009	0.062
140 (2)	0.084	0.090	0.089	0.020	0.009	0.107
200 (3)	0.077	0.082	0.080	0.020	0.009	0.115
	Fwd Delay (old Poa→ RP) (sec)					
45 (8)	0.037	0.037	0.037	0.008	0.007	0.061
70 (6)	0.030	0.029	0.029	0.008	0.007	0.060
140 (2)	0.077	0.082	0.081	0.019	0.007	0.099
200 (3)	0.071	0.075	0.074	0.019	0.006	0.114
	Fwd Delay (RP→new Poa) (sec)					
45 (8)	0.035	0.034	0.035	0.008	0.013	0.060
70 (6)	0.029	0.029	0.029	0.007	0.013	0.058
140 (2)	0.078	0.084	0.083	0.020	0.007	0.108
200 (3)	0.068	0.072	0.071	0.020	0.003	0.124
	Total one-way e2e delay (CN→old PoA→RP→new PoA)					
45 (8)	0.098	0.097	0.098	0.018	0.043	0.163
70 (6)	0.118	0.116	0.117	0.015	0.086	0.171
140 (2)	0.229	0.245	0.244	0.052	0.032	0.279
200 (3)	0.235	0.250	0.247	0.058	0.032	0.322
	Total direct one-way e2e delay (CN→ old/new PoA)					
45 (8)	0.045	0.044	0.045	0.006	0.031	0.068
70 (6)	0.038	0.037	0.037	0.008	0.017	0.069
140 (2)	0.074	0.078	0.077	0.015	0.006	0.093
200 (3)	0.095	0.102	0.100	0.026	0.006	0.144
	Late arrival of MN and delay accumulation in pkt buffering					
45 (8)	0.019	0.018	0.019	0.006	0.007	0.036
70 (6)	0.022	0.023	0.023	0.005	0.009	0.044
140 (2)	0.001	0.001	0.001	0.001	0.001	0.001
200 (3)	0.006	0.006	0.002	0.003	0.009	0.004
	Early MN arrival and delay accumulation in (MN) waiting					
45 (8)	0.026	0.024	0.025	0.009	0.011	0.055
70 (6)	0.027	0.027	0.027	0.007	0.011	0.046
140 (2)	0.105	0.105	0.105	0.011	0.076	0.136
200 (3)	0.088	0.088	0.088	0.015	0.037	0.123

Table F.3: Moments of location (Central tendency) of delay measures observed during HandoffCast forwarding, as function of PoA density and RP node-degree

PoA density (ND)	Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
#	Persistent Handoff Delay (sec)					
45 (8)	0.073	0.085	0.092	0.096	0.106	0.114
70 (6)	0.069	0.079	0.082	0.085	0.094	0.097
140 (2)	0.159	0.175	0.180	0.182	0.186	0.187
200 (3)	0.135	0.160	0.169	0.176	0.188	0.202
	L2-handoff Delay Component (sec)					
45 (8)	0.055	0.065	0.069	0.071	0.075	0.075
70 (6)	0.054	0.063	0.067	0.071	0.075	0.087
140 (2)	0.057	0.068	0.071	0.076	0.088	0.095
200 (3)	0.056	0.071	0.077	0.084	0.101	0.112
	Forwarding Delay Component (old→new PoA) (sec)					
45 (8)	0.035	0.045	0.050	0.054	0.062	0.067
70 (6)	0.027	0.037	0.043	0.047	0.056	0.062
140 (2)	0.084	0.095	0.098	0.100	0.104	0.107
200 (3)	0.071	0.089	0.095	0.101	0.108	0.115
	Fwd Delay (old Poa→RP) (sec)					
45 (8)	0.032	0.041	0.047	0.051	0.060	0.061
70 (6)	0.024	0.035	0.040	0.044	0.053	0.060
140 (2)	0.076	0.087	0.090	0.092	0.095	0.099
200 (3)	0.065	0.083	0.089	0.094	0.102	0.114
	Fwd Delay (RP→new Poa) (sec)					
45 (8)	0.030	0.040	0.047	0.051	0.057	0.057
70 (6)	0.025	0.034	0.039	0.041	0.051	0.055
140 (2)	0.079	0.089	0.092	0.094	0.104	0.106
200 (3)	0.062	0.081	0.088	0.093	0.105	0.12
	Total one-way e2e delay (CN→old PoA→RP→new PoA)					
45 (8)	0.085	0.109	0.121	0.126	0.144	0.163
70 (6)	0.106	0.127	0.140	0.146	0.160	0.171
140 (2)	0.233	0.256	0.262	0.265	0.272	0.279
200 (3)	0.224	0.270	0.284	0.292	0.305	0.322
	Total direct one-way e2e delay (CN→old/new PoA)					
45 (8)	0.040	0.049	0.055	0.057	0.063	0.068
70 (6)	0.032	0.042	0.048	0.052	0.061	0.069
140 (2)	0.072	0.082	0.086	0.088	0.091	0.093
200 (3)	0.086	0.112	0.121	0.127	0.138	0.144
	Late arrival of MN and delay accumulation in pkt buffering					
45 (8)	0.015	0.022	0.028	0.033	0.035	0.036
70 (6)	0.019	0.026	0.028	0.029	0.039	0.044
140 (2)	0.001	0.001	0.001	0.001	0.001	0.001
200 (3)	0.004	0.009	0.009	0.009	0.009	0.009
	Early MN arrival and delay accumulation in (MN) waiting					
45 (8)	0.019	0.032	0.039	0.042	0.051	0.055
70 (6)	0.021	0.032	0.037	0.040	0.045	0.046
140 (2)	0.096	0.114	0.119	0.124	0.130	0.136
200 (3)	0.079	0.097	0.109	0.113	0.120	0.123

Table F.4: Percentiles of delay measures observed during HandoffCast forwarding, as function of PoA density and RP node-degree

PoA density	mean	median	trimean	Std.Dev	Min	Max
#	No. of Probed Channels/R-neighbourhood					
45	4	4	4	0.8721	2	6
70	4	5	4	0.9475	1	8
140	5	5	5	1.1378	1	9
200	5	5	5	1.4446	1	11
	No. of PoA neighbours/R-neighbourhood					
45	5	6	5	1.1963	1	8
70	6	6	6	1.3796	2	10
140	6	6	6	1.6403	2	12
200	7	7	7	2.0256	1	12
	Common PoA neighbours/R-neighbourhood					
45	2	2	2	0.6548	1	4
70	2	2	2	0.6734	1	5
140	2	2	2	0.6737	1	5
200	2	2	2	0.7193	1	5
	Excluded (Pruned) PoA neighbours/R-neighbourhood					
45	3	3	3	0.7138	0	5
70	4	4	4	0.8312	2	7
140	4	4	5	1.2967	1	8
200	5	5	5	1.6875	1	10
	Included (Grafted) PoA neighbours/R-neighbourhood					
45	3	4	3	0.7614	0	5
70	4	4	4	0.9899	1	9
140	4	4	4	1.2627	1	11
200	5	5	5	1.6669	0	11
	PoA node degree					
45 (8)	2	1	2	1.5076	1	8
70 (6)	2	1	2	1.5661	1	12
140 (2)	2	1	2	1.4523	1	10
200 (3)	2	1	2	1.4142	1	9

Table F.5: Moments of central tendency for number of probed channels, R-neighbourhood size, signalling cost (Common,Exclude,Include) neighbours and PoA node-degree, observed during HandoffCast forwarding, as function of PoA density

PoA density	Q_{25}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$Q_{99.9}$
#	No. of Probed Channels/R-neighbourhood					
45	4	5	6	6	6	6
70	4	5	6	6	6	8
140	4	5	6	6	8	9
200	4	6	7	8	9	11
	No. of PoA neighbours/R-neighbourhood					
45	5	6	7	7	8	8
70	5	7	7	8	10	10
140	5	7	8	9	11	12
200	5	8	9	10	12	12
	Common PoA neighbours/R-neighbourhood					
45	1	2	3	3	4	4
70	1	2	3	3	4	5
140	1	2	3	3	4	5
200	1	2	2	3	4	5
	Excluded (Pruned) PoA neighbours/R-neighbourhood					
45	3	4	4	5	5	5
70	3	4	5	5	6	7
140	4	5	6	6	8	8
200	4	6	7	8	9	10
	Included (Grafted) PoA neighbours/R-neighbourhood					
45	3	4	4	4	5	5
70	3	4	5	5	7	9
140	4	5	6	6	8	11
200	4	6	7	8	10	11
	PoA node Degree					
45	1	2	4	6	8	8
70	1	2	4	5	7	12
140	1	3	4	5	8	10
200	1	3	4	5	8	9

Table F.6: Percentiles of probed channels, R-neighbourhood size, signalling cost (Common,Exclude,Include) neighbours and PoA node-degree, observed during HandoffCast forwarding, as function of PoA density

F.7 Persistent Handoff Delay

This section presents the measure of persistent handoff delay by averaging the individual instantaneous delay figures over the entire set of 20 simulation iterations for the particular scenario, as shown in Figures F.5 and F.6.

Furthermore, a regression fit is provided on the observed measure of delay for all 4 PoA densities explored. The grey shadow in each of these graphs tracks the instantaneous measure of variance from the observed mean.

In all four case the fit is acceptable with a residual norm of 0.112, 0.098, 0.137 and

For PoA densities of 140 and 200 AR nodes, the average persistent handoff delay varies between 158 and 142ms respectively. Around 84 and 77ms are owed to the forwarding delay component, while about 62 and 63 ms are attributed to the average observed measure of L2-handoff delay. It may be seen that the forwarding delay component is more than double that of the respective forwarding delay for smaller PoA densities, namely 45 and 70 AR nodes.

From these figures it emerges that: (i) despite the increase in PoA density (which implies a respective increase in the average R-neighbourhood size) the average measure of L2-handoff delay remains approximately constant, (ii) the measure of forwarding delay increases.

With respect to the measure of forwarding delay over the HandoffCast forwarding path segment, Section 5.6.1 has shown that such increase in forwarding delay is not due to the increase on the number of non-leaf ARs withing the topology but as a result of the effective node-degree of the RP node placement.

With respect to the measure of L2-handoff delay, Section 5.6.1 has shown that despite the effective increase in the size of MN's new R-neighbourhood, the number of channels that should be probed within that neighbourhood remains on average *constant*. As a result, by guiding the active AP scanning process over the correct set of channels to be probed, the L2-handoff process of 802.11 can achieve a significantly smaller delay footprint within the measure of persistent handoff delay.

F.8 HandoffCast forwarding path delay components

This section presents the empirical probability density function of one-way delay experienced in the two path segments employs during HandoffCast forwarding.

These are accompanied by the respective measure of forwarding path delay as a function of the number of non-leaf AR within the network topology, at a 95% confidence interval.

We may observe that PoA topologies where the RP is places at an AR with a high node-degree, the observed measure of one-way delay appears to be normally distributed around the mean.

This is also attested from Table F.3 where, for PoA=45 and 70 the means is nearly identical the median delay value (50% percentile). This implies no skew away from the mean value. This is not exactly the case for PoA=140 and 200 which appears to be slightly right-skewed.

then all that is required, is to identify the coordinates of the points defining the length of line segment bc. This may be found easily by identifying the coordinates of points of intersection for each pair of the 3 overlapping circles. For instance, the length of bc is tracked by the points b and c, each found as the top point of intersection between circles CA_A and CA_C for point b, as well as CA_A and CA_B for point c. The third point a is found as the top point of intersection between disks CA_B and CA_C.

We note that the points of intersection between two circles with centre $A(x_1, y_1)$ and $B(x_2, y_2)$ [447] is given by:

$$x_i = \frac{x_2 + x_1}{2} + \frac{(x_2 - x_1)(r_1^2 - r_2^2)}{2d^2} \pm \frac{y_2 - y_1}{2d^2}\sqrt{((r_1 + r_2)^2 - d^2)(d^2 - (r_2 - r_1)^2)} \tag{G.18}$$

$$y_i = \frac{y_2 + y_1}{2} + \frac{(y_2 - y_1)(r_1^2 - r_2^2)}{2d^2} \mp \frac{x_2 - x_1}{2d^2}\sqrt{((r_1 + r_2)^2 - d^2)(d^2 - (r_2 - r_1)^2)} \tag{G.19}$$

where r_1 and r_2 is the radii of A and B respectively, while d is the distance between the centres of two overlapping disks, found in similar fashion by applying the Pythagorean theorem between intersection points (x_i, y_i). We can now iterate the above calculations to identify the circular segment of CA_B defined by chord ca ass well as the circular segment of CA_C defined by chord ab. In a similar fashion we may derive ca and ab as

$$ca = \sqrt{(x_b - x_c)^2 + (y_b - y_c)^2} \tag{G.20}$$

$$ab = \sqrt{(x_c - x_a)^2 + (y_c - y_a)^2} \tag{G.21}$$

Thus we can now compute the total area of overlap between three coverage areas of *different* radii as the sum of 3 circular segments and a straight edge triangle, namely:

$$A(abc) = A_{CircSegA} + A_{CircSegB} + A_{CircSegC} + A_{\triangle abc} \tag{G.22}$$

The area of the third triangle may be found by employing *Heron's formula*, namely:

$$A(\triangle abc) = \sqrt{s(s - ab)(s - bc)(s - ac)} \tag{G.23}$$

where $s(\triangle)$ is the semi-perimeter of the triangle defined as half its perimeter, i.e.

$$s(\triangle abc) = \frac{1}{2}(ab + bc + ac) \tag{G.24}$$

Bibliography

[1] J. Lee, R. Fano, A. Scherr, F. Corbato, and V. Vyssotsky, "Project MAC [Time-sharing Computing Project]," *IEEE Annals of the History of Computing,*, vol. 14, no. 2, pp. 9 – 13, May 1992.

[2] F. J. Corbato and V. A. Vyssotsky, "Introduction and Overview of the Multics System," in *Proceedings of AFIPS 1965 Fall Joint Computer Conference*, Sep. 1965.

[3] J. C. R. Licklider and R. W. Taylor, "The Computer as a Communication Device," *Science and Technology*, vol. 5, no. 76, p. 21, Apr. 1968.

[4] V. Cerf and P. Kirstein, "Issues in packet-network interconnection," *Proceedings of the IEEE*, vol. 66, no. 11, pp. 1386–1408, Nov. 1978.

[5] L. Roberts, "The ARPANET and Computer Networks," *Proceedings of the ACM Conference on The history of personal workstations*, Jan. 1986.

[6] C. Kita, "J.C.R. Licklider's vision for the IPTO," *IEEE Annals of the History of Computing*, vol. 25, no. 3, pp. 62 – 77, 2003.

[7] B. Leiner, D. Cerf, V. Clark, R. Kahn, L. Kleinrock, D. Lynch, J. Postel, L. Roberts, and S. Wolff, "The Past and Future History of the Internet," *Communications of the ACM*, vol. 40, no. 2, pp. 102–108, Feb. 1997.

[8] R. Rutledge, A. Vareha, L. Varian, A. Weis, S. Seroussi, J. Mayer, J. Jaffe, and M. Angell, "An Interactive Network of Time-sharing Computers," in *Proceedings of 24th national conference ACM/CSC-ER*, Aug. 1969.

[9] E. L. Glaser, J. F. Couleur, and G. A. Oliver, "System Design of a Computer for Time Sharing Applications," in *Proceedings of AFIPS 1965 Fall Joint Computer Conference*, Sep. 1965.

[10] F. K. Morioka and R. M. Wainwright, "On-line Communications and the Computer," in *Proceedings of the first ACM symposium on Problems in the optimization of data communications systems*, Oct. 1969.

[11] M. Ohta, "IETF and Internet standards," *IEEE Communications Magazine*, vol. 36, no. 9, pp. 126 - 129, Sep. 1998.

[12] J. Postel, "DOD standard Internet Protocol," *ACM Computer Communication Review*, vol. 10, no. 4, pp. 12 – 51, Oct. 1980.

[13] J. Licklider, "Man-machine Symbiosis," *IRE Transactions on Human Factors in Electronics HFE-1*, vol. 1, no. 3, pp. 4–11, 1960.

[14] I. Jacobs, R. Binder, and E. Hoversten, "General Purpose Packet Satellite Network," *Proceedings of the IEEE*, vol. 66, no. 11, pp. 1448– 1467, Nov. 1978.

[15] R. Kahn, S. Gronemeyer, J. Burchfiel, and R. Kunzelman, "Advances in Packet Radio Technology," *Proceedings of the IEEE*, vol. 66, no. 11, pp. 1468–1497, Nov. 1978.

[16] M. Mouly, M. B. Pautet, and T. Haug, *The GSM System for Mobile Communications.* Telecom Publishing, 1992, ISBN 2-9507190-0-7.

[17] . WG, *IEEE 802.11b High Rate Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification Standard*, IEEE, Oct. 1999.

[18] J. Kruys, "HIPERLAN: Applications and Requirements," in *Proceedings of 3rd IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Oct. 1992.

[19] N. Bambos, "Toward Power-sensitive Network Architectures in Wireless Communications: concepts, issues, and design aspects," *IEEE Personal Communications*, vol. 5, no. 3, pp. 50 – 59, Jun. 1998.

[20] M. Miles, "Recent Advances in lithium Battery Technology," in *23rd Annual Technical Digest on Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, Oct. 2001.

[21] B. Forum, "The Bluetooth Specification Section," Sep. 2001, http://www.bluetooth.com/developer/specification/specification.asp.

[22] D. Egan, "The Emergence of ZigBee in Building Automation and Industrial Control," *Computing & Control Engineering Journal*, vol. 16, no. 2, pp. 14 – 19, May 2005.

[23] G. Abowd, L. Iftode, and H. Mitchell, "The Smart Phone–A First Platform for Pervasive Computing," *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 18 – 19, Jan. 2005.

[24] ITU-T, *Series G: Transmission System and Media, Digital Systems and Networks: One-way transmission time*, May 2000.

[25] R. Steinmetz, "Human perception of jitter and media synchronization," *IEEE Journal in Selected Areas Communications*, vol. 14, no. 1, pp. 61–72, Jan. 1996.

[26] C. Rodbro, M. Christensen, S. Andersen, and S. Jensen, "Compressed domain packet loss Concealment of sinusoidally Coded Speech," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2003.

[27] M.-K. Lee, S.-K. Jung, H.-G. Kang, Y.-C. Park, and D.-H. Youn, "A Packet Loss Concealment algorithm based on Time-Scale Modification for CELP-type speech Coders," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2003.

[28] M. Akimoto and K. Watanabe, "Study on rain attenuation considering rainfall-intensity dependent spatial correlation Characteristics," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, Dec. 2004.

[29] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. New Jersey, US: Printice Hall, 1996.

[30] H. Soliman, C. Castelluccia, K. Malki, and L. Bellier, "Hierarchical Mobile IPv6 mobility management (HMIPv6)," Internet Draft, Dec. 2004, draft-ietf-mipshop-hmipv6-04.txt.

[31] C. Perkins and Ed., "IP Mobility Support," Internet Engineering Task Force, RFC 2002, Oct. 1996.

[32] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6," Internet Engineering Task Force, RFC 3775, June 2004.

[33] R. E. Koodli, "Fast Handovers for Mobile IPv6," Internet Draft, Oct. 2004, draft-ietf-mipshop-fast-mipv6-03.txt.

[34] R. G.114, "One-way Transmission Time," ITU-T, Tech. Rep., May 2000.

[35] G. Karlsson, "Quality Requirements for Multimedia Network Services," in *Proceedings of Radiovetenskap ach kommunikation*, Jun. 1996, pp. 96–100.

[36] R. G.113, "Transmission Impairments due to Speech Processing," ITU-T, Tech. Rep., Feb. 2001.

[37] R. G.711, "Pulse Code modulation (PCM) of Voice Frequencies," ITU-T, Tech. Rep., Nov. 1988.

[38] R. G.726, "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)," ITU-T, Tech. Rep., Apr. 1990.

[39] R. G.723.1, "Dual Rate Speech coder for multimedia communications transmitting at 5.3 and 6.3 kbits/s," ITU-T, Tech. Rep., Mar. 1996.

[40] R. G.729, "Coding of Speech at 8 kbps using Conjugate-structure Algebraic-code-excited Linear-prediction (CS-ACELP)," ITU-T, Tech. Rep., May 1996.

[41] S. Cheshire, "Latency and the Quest for Interactivity," in *White paper commissioned by Volpe Welty Asset Management, L.L.C., for the Synchronous Person-to-Person Interactive Computing Environments Meeting*, Nov. 1996.

[42] H. Balakrishnan, S. Seshan, and R. H. Katz, "Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks," *ACM Wireless Networks*, vol. 1, no. 4, 1995. [Online]. Available: citeseer.nj.nec.com/balakrishnan95improving.html

[43] H. Elaarag, "Improving TCP Performance over Mobile Networks," *ACM Computing Surveys (CSUR)*, vol. 34, no. 3, pp. 357–374, Sep. 2002.

[44] T. Pagtzis and C. Perkins, "Performance Issues for Localised IP Mobility Management," in *Proceedings of IEEE International Conference on Networks (ICON)*, Aug. 2002.

[45] C. Williams, (ed.), H. Soliman, E. Nordmark, T. Pagtzis, J. Kempf, J. Malinen, *et al.*, "Requirements for Localised IP Mobility Management," Internet Draft (Work in progress), Jun. 2001, draft-ietf-mobileip-lmm-requirements-03.txt.

[46] T. Pagtzis, C. Williams, C. Perkins, and P. Kirstein, "Requirements for Localised IP Mobility Management," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Louisiana, New Orleans, Mar. 2003.

[47] T. Pagtzis and P. Kirstein, "Proactive Mobility for Future IP Wireless Access Networks," in *Proceedings of 6th IASTED/IEEE International Conference on Wireless and Optical Communications (WOC)*, Jul. 2002.

[48] T. Pagtzis and P. T. Kirstein, "On the Performance of CFP Repetition Rate in IEEE 802.11 for IP traffic in Mobility-enabled WLANs," in *Proceedings of 27th Annual IEEE Conference on Local Computer Networks (LCN)*, Tampa, Florida, Nov. 2002.

[49] T. Pagtzis, P. Kirstein, S. Hailes, and H. Afifi, "Proactive Seamless Mobility Management for Future IP Radio Access Networks," *Computer Communications*, vol. Special Issue on '3G Wireless and Beyond', pp. 1267–1280, Jul. 2003.

[50] T. Pagtzis and P. Kirstein, "A Model for Proactive Seamless IP Mobility and Mobility-Hop Routing," in *Proceedings of 10th IEEE International Conference on Networks (ICON)*, Singapore, Aug. 2002.

[51] T. Pagtzis, P. Kirstein, and S. Hailes, "Operational and Fairness Issues with Connection-less Traffic over IEEE802.11b," in *Proceedings of IEEE International Conference on Communications (ICC)*, Jun. 2001.

[52] T. Pagtzis, S. Hailes, and P. Kirstein, "The Case for Proactive Mobile IPv6," in *Proceedings of the 28th Annual IEEE Conference on Local Computer Networks (LCN)*, Oct. 2003.

[53] H. Bolcskel, A. Paulraj, K. Hari, R. Nabar, and W. Lu, "Fixed Broadband Wireless Access: State of the Art, Challenges, and Future Directions," *IEEE Personal Communications*, vol. 39, no. 1, pp. 100–108, Jan. 2001.

[54] W. Webb, "Broadband Fixed Wireless Access as a key Component of the Future Integrated Communications Environment," *IEEE Computer Communications*, vol. 39, no. 9, pp. 115–121, Sep. 2001.

[55] A. Jamalipour and T. Tung, "The Role of Satellites in Global IT: Trends and Implications," *IEEE Personal Communications*, vol. 8, no. 3, pp. 5–11, Jun. 2001.

[56] L. Bos and S. Leroy, "Toward an all-IP-based UMTS System Architecture," *IEEE Network*, vol. 15, no. 1, pp. 36–45, Jan. 2001.

[57] J. Huber, D. Weiler, and R. Brand, "UMTS, the Mobile Multimedia Vision for IMT 2000: a Focus on Standardization," *IEEE Personal Communications*, vol. 38, pp. 129–136, Sep. 2000.

[58] A. Fasbender, F. Reichert, E. Geulen, J. Hjelm, and T. Wierlemann, "Any Network, any Terminal, Anywhere," *IEEE Personal Communications*, vol. 6, no. 2, pp. 22–30, Apr. 1999.

[59] T. Farnham, G. Clemo, R. Haines, *et al.*, "IST-TRUST: a Perspective on the Reconfiguration of Future Mobile Terminals using Software Downloads," in *Proceedings of 11th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 2000, pp. 1054–1059.

[60] M. Oliphant, "The Mobile Phone meets the Internet," *IEEE Spectrum*, vol. 36, no. 8, pp. 20–28, Aug. 1999.

[61] R. Want, B. Schilit, N. Adams, R. Gold, K. Petersen, D. Goldbery, M. Ellis, and M. Weiser, *Mobile Computing.* Kluwer Publishing, Feb. 1997, ch. *"The ParcTab Ubiquitous Computing Experiment"*, pp. 45–101.

[62] R. Jana, T. Johnson, S. Muthukrishnan, and T. Vitaletti, "Location-based Services in a Wireless WAN using Cellular Digital Packet Data (CDPD)," in *Proceedings of 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access*, Jul. 2001, pp. 74–80.

[63] Y. Inoue and M. Nakagawa, "MAC protocol for Inter-vehicle Communication Network using Spread-Spectrum Techniques," in *Proceedings of Vehicle Navigation and Information Systems Conference*, 1994, pp. 149–152.

[64] S. Tsugawa, S. Kato, K. Tokuda, T. Matsui, and H. Fujii, "A Cooperative Driving System with Automated Vehicles and Inter-vehicle Communications in Demo 2000," in *Proceedings of Intelligent Transportation Systems*, Jul. 2001, pp. 918 –923.

[65] H. Koshima and J. Hoshen, "Personal Locator Services Emerge," *IEEE Spectrum*, vol. 37, no. 2, pp. 41–48, Feb. 2000.

[66] L. Hanzo, P. Cherriman, and E. L. Kuan, "Interactive Cellular and Cordless Video Telephony: State of the Art System Design Principles and Expected Performance," *Proceedings of the IEEE*, vol. 88, no. 9, pp. 1388 –1413, Sep. 2000.

[67] T. Henderson, "Latency and User-behaviour on a Multiplayer Game Server," in *Proceedings of 3rd International Workshop on Networked Group Communication (NGC)*, Nov. 2001.

[68] G. Newsroom, "3G - Future Terminal Designs," Aug. 2001, http://www.3gnewsroom.com/html/images/index.shtml.

[69] T. La Porta, K. Sabnani, and R. Gitlin, "Challenges for Nomadic Computing: Mobility Management and Wireless Communications," *ACM Journal in Mobile Networking and Applications*, vol. 1, no. 1, pp. 3–16, 1996.

[70] L. Kleinrock, "Nomadic Computing - an Opportunity," *ACM Computer Communication Review*, vol. 25, no. 1, pp. 38–47, Jan. 1995.

[71] W.-F. Alliance, "Wireless Fidelity Alliance," Jan. 2001, http://www.wi-fi.org/OpenSection/index.asp.

[72] 3GPP, "3G Specifications Section," Sep. 2001, http://www.3gpp.org/3G_Specs/3G_Specs.htm.

[73] 3GPP2, "3G (Parallel) Specification Section," Sep. 2001, http://www.3gpp2.org/Public_html/specs/index.cfm.

[74] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 884–895, Jun. 1999.

[75] I. Lee, A. Chan, and C. Sundberg, "Space-Time Bit-interleaved Coded Modulation for OFDM Systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 3, pp. 820 – 825, Mar. 2004.

[76] Y. Cheng, Y. Lee, and T. Sheu, "Multi-rate transmissions in infrastructure wireless lan based on ieee 802.11b protocol," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Fall)*, Oct. 2001.

[77] C. Hansen and G. Pottie, "A Distributed Access Algorithm for Cellular Radio Systems with Channel Partitioning," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 76–82, 1999.

[78] C. Chang, J. Liang, A. Paulraj, and T. Kailath, "A Multiple Access Technique for Cellular Packet Networks with Admission Control," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, Nov. 1994.

[79] A. Veres, A.and Campbell, M. Barry, and L. Sun, "Supporting Service Differentiation in Wireless Packet Networks using Distributed Control," *IEEE Journal in Selected Areas Communications*, vol. 19, no. 10, pp. 2081–2093, Oct. 2001.

[80] W. 802i1, *IEEE 802.11 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification Standard*, IEEE, Oct. 1999.

[81] E. . 113/v6.3.1, *Digital Cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service Description Stage 1/2*, ETSI, Nov. 2001.

[82] F. Chiussi, D. Khotimsky, and S. Krishnan, "Mobility Management in Third Generation all-IP Networks," *IEEE Communications Magazine*, vol. 40, no. 9, pp. 124–135, Sep. 2002.

[83] R. Klukas and M. Fattouche, "Line-of-sight Angle of Arrival Estimation in the Outdoor Multipath Environment," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 1, pp. 342–351, Feb. 1998.

[84] M.-H. Chiu and M. Bassiouni, "Predictive schemes for handoff prioritization in cellular networks based on mobile positioning," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 510–522, Mar. 2000.

[85] M. Akar and U. Mitra, "Joint Power and Handoff Control using a Hybrid Systems Framework," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2003.

[86] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. Makela, R. Pichna, and J. Vallstron, "Handoff in Hybrid Mobile Data Networks," *IEEE Personal Communications*, vol. 7, no. 2, pp. 34–47, Apr. 2000.

[87] I. Akyildiz and J. Ho, "A Mobile User Location Update and Paging Mechanism under Delay Constraints," *ACM Computer Communication Review*, vol. 25, no. 4, pp. 244 – 255, Oct. 1995.

[88] R. Gau and Z. Haas, "Concurrent Search of Mobile users in Cellular Networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 1, pp. 117–130, Feb. 2004.

[89] J. Ioannidis, D. Duchamp, and G. Maguire, "IP-based Protocols for Mobile Internetworking," in *Proceedings of ACM SIGCOMM*, Sep. 1991, p. 235245.

[90] J. Ioannidis and G. Maguire, "The Design and Implementation of a Mobile Internet-working Architecture," in *Proceedings of Winter USENIX Symposium on Mobile and Location Independent Computing*, San Diego, CA, Jan. 1993, pp. 491–502.

[91] Y. Rekhter and C. Perkins, "Optimal routing for mobile hosts using ip's loose source route option," Internet-Draft, Oct. 1992.

[92] C. Perkins, A. Myles, and D. Johnson, "IMHP: a Mobile Host Protocol for the Internet," in *Proceedings of conference on Internet Society/5th joint European networking conference*, Dec. 1994.

[93] P. Tasman, "Protocols and Caching Strategies in Support of Internetwork Mobility," University of Wisconsin at Madison, Madison, WI, Tech. Rep., 1994.

[94] F. Teraoka and M. Tokoro, "Host Migration Transparency in IP Networks: The VIP Approach," *ACM Computer Communication Review*, vol. 23, no. 1, pp. 45–65, Jan. 1993.

[95] T. Blackwell *et al.*, "Secure Short-Cut Routing for Mobile IP," in *Proceedings of Summer USENIX Symposium on Mobile and Location Independent Computing*, Jun. 1994.

[96] R. Hager, A. Klemets, G. Maguire, M. Smith, and F. Reichert, "Mint - a mobile internet router," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Spring)*, May 1993.

[97] V. Gupta and A. Dixit, "The Design and Deployment of a Mobility Supporting Network," in *Proceedings of International Symposium on Parallel Architectures, Algorithms, and Networks*, Jun. 1996.

[98] P. Bhagwat and C. Perkins, "A mobile network system based on internet protocol (ip)," in *Proceedings of USENIX Symposium on Mobile and Location Independent Computing*, Cambridge, MA, Aug. 1993.

[99] A. Myles and D. Skellern, "Comparing Four IP-based Mobile Host Protocols," *Computer Networks and ISDN Systems*, vol. 26, no. 3, pp. 349–355, Nov. 1993.

[100] S. Cheshire and M. Baker, "Internet mobility 4x4," *ACM Computer Communication Review*, vol. 26, no. 4, pp. 318–329, Oct. 1996.

[101] C. Perkins and Ed., "IP Mobility Support for IPv4," Internet Engineering Task Force, RFC 3220, Jan. 2002.

[102] ——, "IP Mobility Support for IPv4," Internet Engineering Task Force, RFC 3344, Aug. 2002.

[103] S. Deering, "ICMP Router Discovery Messages," Internet Engineering Task Force, RFC 1256, Sep. 1991.

[104] R. Caceres and L. Iftode, "The Effects of Mobility on Reliable Transport Protocols," in *Proceedings of 14th International Conference on Distributed Computing Systems*, Jul. 1994, pp. 12–20.

[105] H. Yokota, A. Idoue, T. Hasegawa, and T. Kato, "Link layer assisted mobile ip fast handoff method over wireless lan networks," in *Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, US, Sep. 2002.

[106] S. Deering and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," Internet Engineering Task Force, RFC 2460, Dec. 1998.

[107] T. Narten, E. Nordmark, and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)," Internet Engineering Task Force, RFC 2461, Dec. 1998.

[108] A. Conta and S. Deering, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification," Internet Engineering Task Force, RFC 2463, Dec. 1998.

[109] M. Crawford, "Transmission of IPv6 Packets over Ethernet Networks," Internet Engineering Task Force, RFC 2464, Dec. 1998.

[110] S. Thomson and T. Narten, "IPv6 Stateless Address Autoconfiguration," Internet Engineering Task Force, RFC 2462, Dec. 1998.

[111] C. Perkins and D. Johnson, "Mobility support in IPv6," in *Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, Nov. 1996.

[112] R. Droms, "Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6," Internet Engineering Task Force, RFC 3736, Apr. 2004.

[113] G. Montenegro and Ed., "Reverse Tunneling for Mobile IP, revised," Internet Engineering Task Force, RFC 3024, Jan. 2001.

[114] R. Moskowitz, "Host Identity Protocol," Internet Draft, Feb. 2005, draft-ietf-hip-base-02.txt.

[115] T. Henderson, J. Ahrenholz, and J. Kim, "Experience with the Host Identity Protocol for Secure Host Mobility and Multihoming," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Jun. 2003.

[116] L. Ong and J. Yoakum, "An Introduction to the Stream Control Transmission Protocol (SCTP)," Internet Engineering Task Force, RFC 3286, May 2002.

[117] M. Matsumoto, M.and Kozuka, K. Fujikawa, and O. Y., Internet Draft, Oct. 2003, http://www.ietf.org/internet-drafts/draft-arifumi-tcp-mh-00.txt.

[118] H. Schulzrinne and E. Wedlund, "Application-layer Mobility using SIP," *ACM Computer Communication Review*, vol. 4, no. 3, pp. 47–57, Jul. 2000.

[119] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: Session Initiation Protocol," Internet Engineering Task Force, RFC 2543, Mar. 1999.

[120] J. Klensin, "Role of the Domain Name System (DNS)," Internet Engineering Task Force, RFC 3467, Feb. 2003.

[121] V. PadManaphan and R. Katz, "Using dns to support mobility," University of California, Berkley, http://research.microsoft.com/ padmanab/talks/ilp96.ps, Tech. Rep., Mar. 1996.

[122] A. Pappas, "Research Proposal: Real-Time DNS," Jan. 2002, http://www.ee.ucl.ac.uk/ apappas/Proposal.htm.

[123] A. McAuley and R. Morera, "Name and Address Decoupling in Support of Dynamic Networks," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, Oct. 2002.

[124] H.-I. Liu, "Mobile Domain Name System: an Alternative for Mobile IP," in *Proceedings of The 8th International Conference on Communication Systems*, Nov. 2002.

[125] R. Austein, "Tradeoffs in Domain Name System (DNS) Support for Internet Protocol version 6 (IPv6)," Internet Engineering Task Force, RFC 3364, Aug. 2002.

[126] M. Y. L. Junginger, "A Self-organizing Publish/Subscribe Middleware for Dynamic Peer-to-Peer Networks," *IEEE Network*, vol. 18, no. 1, pp. 38–43, Feb. 2004.

[127] I. Stoica, R. Morris, D. Karger, M. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," in *Proceedings of Annual Conference on Applications, technologies, architectures, and protocols for computer communications*, Aug. 2001.

[128] I. Stoica, D. Adkins, S. Zhuang, S. Shenker, and S. Surana, "Internet indirection infrastructure," in *Proceedings of Annual Conference on Applications, technologies, architectures, and protocols for computer communications*, Aug. 2002.

[129] T. Pagtzis, "A Framework for Proactive Mobility in Mobile IPv6," Internet Draft (Work in progress), Jul. 2001, draft-pagtzis-mobileip-proactivev6-00.txt.

[130] A. Campbell, J. Gomez, S. Kim, A. Valko, C.-Y. Wan, and Z. Turanyi, "Design, Implementation, and Evaluation of Cellular-IP," *IEEE Personal Communications*, vol. 7, no. 4, pp. 42–49, Aug. 2000.

[131] J. Jubin and J. D. Tornow, "The DARPA Packet Radio Network Protocol," *Proceedings of IEEE*, vol. 75, pp. 21–32, 1987.

[132] T. Kurita, S. Iai, and N. Kitawaki, "Effects of Transmission Delay in Audiovisual Communication," *Electronics and Communications in Japan*, vol. 77, no. 3, pp. 63–74, 1995.

[133] E. T. WG5, "Voice Quality Recommendations for IP Telephony," ETSI," PN-4689, Feb. 2000.

[134] Y. Wang, M. Claypool, and Z. Zuo, "An Empirical Study of RealVideo Performance Across the Internet," in *Proceedings of ACM SIGCOMM Internet Measurement Workshop*, Nov. 2001.

[135] I. Kouvelas, V. Hardman, and A. Watson, "Lip-synchronisation for use over the Internet: Analysis and Implementation," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, Nov. 1996.

[136] T. Barnes and M. Smith, "The Effects of Visual Discrimination of Image Movements across the Retina," *Aviation Space and Environmental Medicine*, vol. 52, no. 8, p. 466472, 1981.

[137] S. Handel, *Listening: An Introduction to the Perception of Auditory Events*, 3rd ed. Cambridge, MA: MIT Press, 1989.

[138] B. Maki and W. Mcllroy, "The Control of Foot Placement during Compensatory Stepping Reactions: Does Speed of Response take Precedence over Stability?" *IEEE Transactions on Rehabilitation Engineering*, vol. 7, no. 1, pp. 80–90, Mar. 1999.

[139] R. W. Bailey, *Human Performance Engineering - Using Human Factors/Ergonomics to Achieve Computer System Usability*, 2nd ed. Prentice Hall, Englewood Cliffs, 1989.

[140] I. K. MacKenzie and C. Ware, "Lag as a Determinant of Human Performance in Interactive Systems," in *Proceedings of CHI '93 Conference on Human factors in computing systems*, Apr. 1993, pp. 488–493.

[141] K. S. Park and R. S. Kenyon, "Effects of Network Characteristics on Human Performance in a Collaborative Virtual Environment," in *Proceedings of IEEE Virtual Reality*, Mar. 1999, pp. 104–111.

[142] R. Becher, M. Dillinger, M. Haardt, and W. Mohr, "Broadband Wireless Access and Future Communication Networks," *Proceedings of the IEEE*, vol. 89, no. 1, pp. 58–75, Jan. 2001.

[143] S. Vaughan-Nichols, "The challenge of Wi-Fi Roaming," *Computer*, vol. 36, no. 7, pp. 17–19, Jul. 2003.

[144] J. Pereira, "Fourth generation: now, it is personal!" in *Proceedings of 11thIEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2000, pp. 1009–1016.

[145] T. Siep, I. Gifford, R. Braley, and R. Heile, "Paving the Way for Personal Area Network Standards: an Overview of the IEEE P802.15 Working Group for Wireless Personal Area Networks," *IEEE Personal Communications*, vol. 7, no. 1, pp. 37–43, Feb. 2000.

[146] G. P. Pollini, S. Meier-Hellstern, and D. J. Goodman, "Signaling Traffic Volume Generated by Mobile and Personal Communications," *IEEE Personal Communications*, pp. 60–65, Jun. 1995.

[147] R. Chalmers, G. Krishnamurthi, and K. Almeroth, "Enabling Intelligent Handovers in Heterogeneous Wireless Networks," *ACM Journal in Mobile Networking and Applications*, to appear 2005.

[148] R. Katz, "Adaptation and Mobility in Wireless Information Systems," *IEEE Communications Magazine*, vol. 40, no. 5, pp. 102–114, May 2002.

[149] M. Stemm and R. Katz, "Vertical handoffs in wireless overlay networks," *ACM Journal in Mobile Networking and Applications*, vol. 3, no. 4, pp. 335–350, Dec. 1998.

[150] D. Avidor, C. Hartmann, and S. Mukherjee, "On the Uplink SIR of a Mobile in Soft Handoff with Applications to BLER Estimation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 467–474, Mar. 2005.

[151] B. Cheung and V. Leung, "Network Configurations for Seamless Support of CDMA Soft Handoffs between Cell Clusters," *IEEE Journal in Selected Areas Communications*, vol. 15, no. 7, pp. 1276 – 1288, Sep. 1997.

[152] J. Wu, S. Affes, and P. Mermelstein, "Forward-link Soft-handoff in CDMA with Multiple-antenna Selection and fast joint Power Control," *IEEE Transactions on Wireless Communications*, vol. 2, no. 3, pp. 459–471, May 2003.

[153] C. V. Bovy *et al.*, "Analysis of End-to-End Delay Measurements in the Internet," in *Proceedings of Passive and Active Measurements Workshop (PAM)*, Mar. 2002. [Online]. Available: http://www.ripe.net/projects/ttm/Documents/PAM2002_TUD.pdf

[154] S. Mukkamalla and B. Raman, "Scaling and Latency Issues in Mobile-IP," http://www.cs.berkeley.edu/ adj/cs294-1.s98/projects/MobileIP/sld001.htm, Apr. 1998.

[155] C. M. Project, "Mobile IP Implementations," Web Page, Oct. 2001.

[156] H. Jiang and C. Dovrolis, "Passive Estimation of TCP Round-trip Times," *ACM Computer Communication Review*, vol. 32, no. 3, pp. 75–88, Jul. 2002.

[157] R. Carter and M. Crovella, "Server Selection using Dynamic Path Characterization in Wide-area Networks," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Apr. 1997.

[158] J. Gwertzman and M. Seltzer, "The Case for Geographical Push-Caching," in *Proceedings of Fifth Workshop on Hot Topics in Operating Systems*, May 1995.

[159] A. Acharya and J. Saltz, "A Study of Internet Round-trip Delay," in *Technical Report CS-TR-3736*, University of Maryland, Dec. 1996.

[160] J. Guyton and M. Schwartz, "Locating Nearby Copies of Replicated Internet Servers," *ACM Computer Communication Review*, vol. 25, no. 4, pp. 288–298, Oct. 1995.

[161] W. F. Lo and H. T. Mouftah, "Carrier sense multiple access with collsision detection for radio channels," *IEEE 13th International Communications and Energy Conference*, pp. 244–257, Apr. 1984.

[162] C. Chen, "Medium Access Control of Wireless LANs for Mobile Computing," *IEEE Network*, vol. 8, no. 5, pp. 50–63, Mar. 1994.

[163] P. Karn, "MACA - a New Channel Access Method for Packet Radio," in *9th Computer Networking Conference in ARRL/CRRL Amateur Radio*, Nov. 1990, pp. 13 440–13 450.

[164] C. Fullmer and J. Garcia-Luna-Aceves, "FAMA-PJ: a Channel Access Protocol for Wireless LANs," in *Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, Dec. 1995.

[165] L. Kleinrock and M. . Scholl, "Packet Switching in Radio Channels: New Conflict-Free Multiple Access Schemes," *IEEE Transactions on Communications*, vol. 28, pp. 1015–1029, Aug. 1980.

[166] L. Kleinrock and F. A. Tobagi, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple-access Modes and their Throughput-Delay Characteristics," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1400–1416, Jun. 1975.

[167] A. Colvin, "CSMA with Collision Avoidance," *IEEE Computer Communications*, vol. 6, no. 5, pp. 227–235, Jul. 1983.

[168] A. Snoeren, H. Balakrishnan, and M. Kaashoek, "Reconsidering Internet Mobility," in *Proceedings of 8th Workshop on Hot Topics in Operating Systems (HotOS-VIII)*, May 2001.

[169] J. Finney and A. Scott, "Implementing Mobile IPv6 for Multimedia," in *Proceedings of GEMISIS/IEE/BCS Symposium on Multimedia Network Technology*, vol. Digest No. G/MNT/1/1998, May 1998.

[170] J. Vatn, "Long Random Wait Times for Getting a Care-of Address are a Danger to Mobile Multimedia," in *Proceedings of International Workshop and on Mobile Multimedia Communications (MOMuC)*, Nov. 1999.

[171] J. Vatn and G. Q. Maguire, "The effect of Using co-located care-of Addresses on Macro-handover Latency," in *Proceedings of 14th Nordic Tele-traffic Seminar (NTS 14)*, Aug. 1998.

[172] A. Mcauley, S. Das, S. Baba, *et al.*, "Dynamic Registration and Configuration Protocol (DRCP),"
draft-itsumo-drcp-01.txt, Internet Engineering Task Force, Internet Draft (Work in progress), Jul. 2000. [Online]. Available: "http://www.ietf.org/internet-drafts/ "draft-itsumo-drcp-01.txt

[173] S. Dutta, J. Chen, S. Das, S. Madhani, A. McAuley, *et al.*, "Implementing a Testbed for Mobile Multimedia," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, 2001.

[174] P. Nikander, Ed., J. Kempf, and E. Nordmark, "IPv6 Neighbor Discovery (ND) Trust Models and Threats," Internet Engineering Task Force, RFC 3756, May 2004.

[175] J. Arkko, Ed., J. Kempf, B. Zill, and P. Nikander, "SEcure Neighbor Discovery (SEND)," Internet Engineering Task Force, RFC 3971, Mar. 2005.

[176] S. Faccin, F. Le, B. Patil, C. Perkins, F. Dupont, and M. Maknavicius, "Mobile IPv6 Authentication, Authorization, and Accounting Requirements," draft-le-aaa-mipv6-requirements-02.txt, Internet Engineering Task Force, Internet Draft (Work in progress), Apr. 2003. [Online]. Available: http://tools.ietf.org/wg/mipv6/draft-le-aaa-mipv6-requirements-02.txt

[177] A. Neumann, X. Fu, and H. Karl, "Prototype implementation and performance evaluation of a qos-conditionalized handoff scheme for mobile ipv6 networks," in *Proceedings of IEEE 18th Annual Workshop on Computer Communications*, Oct. 2003.

[178] S. Paskalis, A. Kaloxylos, E. Zervas, and L. Merakos, "An Efficient RSVP-mobile IP Interworking Scheme," *ACM Journal in Mobile Networking and Applications*, vol. 8, no. 3, pp. 197–207, Jun. 2003.

[179] A. Mukhopadhyay and D. Saha, "Performance Analysis of Mobile Wireless Overlay Networks based on Application Services," in *Proceedings of IEEE International Conference on Personal Wireless Communications*, Jan. 2005.

[180] N. Montavont and T. Noel, "Analysis and evaluation of mobile ipv6 handovers over wireless lan," *ACM Journal in Mobile Networking and Applications*, vol. 8, no. 6, pp. 643–653, 2003.

[181] J. Lee, T. Jung, S. Yoon, S. Youm, and C. Kang, "An Adaptive Resource Allocation Mechanism including Fast and Reliable Handoff in IP-based 3G Wireless Networks," *IEEE Personal Communications*, vol. 7, no. 6, pp. 42–47, Dec. 2000.

[182] T. Ernst, C. Castelluccia, and H. Lach, "Extending Mobile-IPv6 with Multicast to support mobile networks in IPv6," in *Proceedings of IEEE European Conference on Universal Multiservice Networks*, Colmar, France, Oct. 2000.

[183] S. Seshan, H. Balakrishnan, and R. Katz, "Handoffs in Cellular Wireless Networks: The Daedalus Implementation and Experience," 1996.

[184] J. Mysore and V. Bharghavan, "A New Multicasting-based Architecture for Internet Host Mobility"," in *Proceedings of 3rd annual ACM/IEEE International conference on Mobile computing and Networking*, Sep. 1997, pp. 161–172.

[185] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, and L. Wei, "Protocol Independent Multicast-Sparse Mode

(PIM-SM): Protocol Specification," Internet Engineering Task Force, RFC 2117, Jul. 1997.

[186] A. Ballardie, "Core Based Trees (CBT ver. 2) Multicast Routing," Internet Engineering Task Force, RFC 2189, Sep. 1997.

[187] D. Waitzman, C. Partridge, and S. Deering, "Distance Vector Multicast Routing Protocol," Internet Engineering Task Force, RFC 1075, Nov. 1988.

[188] Y. Wang and W. Chen, "Supporting IP Multicast for Mobile Hosts," *ACM/Kluwer special issue on wireless internet and intranet access*, pp. 57–66, Sep. 2001.

[189] A. Delgado, A. Mihailovic, N. Georganopoulos, and A. Aghvami, "Evaluating the Performance of Transport Protocols over MMP," in *Proceedings of London Communications Symposium*, 2000.

[190] J. Lee, "SGM support in Mobile-IP," Internet Draft (Work in progress), Oct. 2000.

[191] Y. Ezaki and Y. Imai, "Mobile IPv6 handoff by Explicit Multicast," Internet Engineering Task Force, May 2001.

[192] R. Boivie and N. Feldman, "Small Group Multicast," Internet Draft (Work in progress), Feb. 2001.

[193] A. E. Yegin, M. Parthasarathy, and C. Williams, "Mobile IPv6 Neighborhood Routing for Fast Handoff," draft-yegin-mobileip-nrouting-01.txt, Internet Engineering Task Force, Internet Draft (Work in progress), Oct. 2000. [Online]. Available: http://www.ietf.org/internet-drafts/draft-yegin-mobileip-nrouting-01.tx%t

[194] L. Kleinrock, "Nomadicity: anytime, anywhere in a disconnected world," *Mobile Networks and Applications*, 1996.

[195] B. Xu, S. Hischke, and B. Walke, "The Role of Ad hoc Networking in Future Wireless Communications," in *Proceedings of International Conference on Communication Technology*, Apr. 2003.

[196] J. Gruber, "A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection," *IEEE Transactions on Communications*, vol. 30, no. 6, pp. 728 – 738, Apr. 1982.

[197] K. Sriram and W. Whitt, "Characterising Superposition Arrival Processes in Packet Multiplexers for Voice and Data," *IEEE Journal in Selected Areas Communications*, pp. 833–846, Sep. 1986.

[198] H. Lee and C. Un, "A study of on-off characteristics of conversational speech," *IEEE Transactions on Communications*, vol. 34, no. 6, pp. 630 – 637, Jun. 1986.

[199] P. Brady, "A Technique for Investigating on/off Patterns of Speech," *Bell Labs Tech. Journal*, vol. 44, no. 1, pp. 1–22, Jan. 1965.

[200] ——, "A Model for generating On-Off Speech Patterns in Two-way Conversation," *Bell System Technical Journal*, vol. 48, no. 9, pp. 2445–2472, Sep. 1969.

[201] W. Jiang and H. Schulzrinne, "Analysis of on-off Patterns in VoIP and their Effect on Voice Traffic Aggregation," in *Proceedings of Ninth International Conference on Computer Communications and Networks*, Oct. 2000.

[202] B. Chen and C. Sundberg, "An Integrated Error Correction and Detection System for Digital Audio Broadcasting," *IEEE Transactions on Broadcasting*, vol. 46, no. 1, pp. 68–78, Mar. 2000.

[203] W. Zeng and B. Liu, "Geometric-structure-based Error Concealment with Novel Applications in block-based Low-bit-rate Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 648 – 665, 1999.

[204] C. Perkins, O. Hodson, and V. Hardman, "A Survey of Packet Loss Recovery Techniques for Streaming Audio," *IEEE Network*, vol. 12, no. 5, pp. 40–48, Sep. 1998.

[205] A. Friday, M. Wu, J. Finney, S. Schmid, *et al.*, "Network-layer Access Control for Context-aware IPv6 Applications," *Wireless Networks*, vol. 9, no. 4, Jul. 2003.

[206] H. T. University, "The HUT Mobile IPv6 implementation," http://www.mobile-ipv6.org/, 2003.

[207] B. Pentland, "Fast Router Advertisement for IPv6 Implementation," http://www.ctie.monash.edu.au/ipv6/fastra.htm, 2003.

[208] C. Perkins and O. Hodson, "RAT: Robust Audio Tool," http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/, 2001.

[209] B. Biggs *et al.*, "KPhone: A Voice over Internet phone," http://www.wirlab.net/kphone/, 2004.

[210] G. Anastasi, E. Borgia, and E. Conti, M.and Gregori, "Wi-Fi in Ad-hoc Mode: a Measurement Study," in *Proceedings of Pervasive Computing and Communications*, May 2004.

[211] C. Perkins, "IP Encapsulation within IP," Internet Engineering Task Force, RFC 2003, Oct. 1996.

[212] E. Parzen, *Modern Probability Theory and Its Applications*, 2nd ed. New York, NY: John Wiley, 1960.

[213] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York, NY: John Wiley, 1968.

[214] S. Bohacek, J. Hespanha, J. Lee, and K. Obraczka, "A Hybrid Dystems Modeling Gramework for Fast and Accurate Simulation of Data Communication Networks," in *Proceedings of ACM International Conference on Measurement and Modeling of Computer Systems*, Jul. 2003, pp. 58–69.

[215] A. Mukherjee, "On the Dynamics and Significance of Low Frequency Components of Internet Load," *Internetworking: Research and Experience*, vol. 5, pp. 163–205, Dec. 1994.

[216] V. Paxson, "End-to-End Internet Packet Dynamics," in *Proceedings of ACM SIGCOMM*, Sep. 1997.

[217] A. Markopoulou, "Assessing the Quality of Multimedia Communications over Internet Backbone Networks," Ph.D. dissertation, Stanford University, Oct. 2002.

[218] R. Cole and J. Rosenbluth, "Voice over IP Performance Monitoring," *ACM Computer Communication Review*, vol. 31, no. 2, Apr. 2001.

[219] A. Markopoulou, F. Tobagi, and M. Karam, "Assessing the Quality of Voice Communications over Internet Backbones," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, Oct. 2003.

[220] T. Ezaki, H. Kawakami, and K. Asatani, "A new voip call admission control scheme with use of alternate routing for low call loss probability," in *Proceedings of IEEE International Conference on Communications (ICC)*, Jun. 2004.

[221] Y. Ito, S. Tasaka, and Y. Fukuta, "Psychometric Analysis of the Effect of End-to-End Delay on User-level QoS in Live Audio-Video Transmission," in *Proceedings of IEEE International Conference on Communications (ICC)*, Jun. 2004.

[222] B. Schilit, A. LaMarca, G. Borriello, *et al.*, "Vision & challenges: Challenge: Ubiquitous Location-aware Computing and the "Place-Lab" Initiative," in *Proceedings of 1st ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, Sep. 2003.

[223] I. Clarke, S. Miller, T. Hong, O. Sandberg, and B. Wiley, "Protecting Free Expression Online with Freenet," *IEEE Internet Computing*, vol. 6, no. 1, pp. 40–49, Jan. 2002.

[224] A. Fitkov-Norris, E.D.; Khanifar, "Dynamic pricing in mobile communication systems," in *Proceedings of 1st International Conference on 3G Mobile Communication Technologies*, Mar. 2000.

[225] K. Tachikawa, "A Perspective on the Evolution of mobile communications," *IEEE Communications Magazine*, vol. 41, no. 10, pp. 66–73, Oct. 2003.

[226] S. Arbanowski, P. Ballon, K. David, O. Droegehorn, H. Eertink, *et al.*, "I-centric communications: Personalization, Ambient Awareness, and Adaptability for future Mobile Services," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 63–69, Sep. 2004.

[227] H. Wang, R. Katz, and J. Giese, "Policy-enabled Handoffs across Heterogeneous Wireless Networks," in *Proceedings of 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, 1999.

[228] A. Valko, "Cellular IP: A New Approach to Internet Host Mobility," *ACM Computer Communication Review*, vol. 29, no. 3, pp. 55–65, Jan. 1999. [Online]. Available: theboringoldpaperforcellIPtogiveasref

[229] R. Ramjee, K. Varadhan, L. Salgarelli, S. R., *et al.*, "HAWAII: a Domain-based Approach for supporting Mobility in Wide-area Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, pp. 396–410, Jun. 2002.

[230] C. Perkins, J. Malinen, and F. Le, "Mobile IPv6 Regional Registrations," Internet Draft (Work in progress), Jul. 2001, draft-malinen-mobileip-regreg6-02.txt.

[231] A. Das, S. and Mcauley, "IDMP: an Intra-domain Mobility Management Protocol for Next-generation Wireless Networks," *IEEE Wireless Communications*, vol. 9, no. 3, pp. 130–138, June 2002.

[232] A. Helmy, M. Jaseemuddin, and G. Bhaskara, "Multicast-based Mobility: a Novel Architecture for Efficient Micr-mobility," *IEEE Journal in Selected Areas Communications*, vol. 22, no. 4, pp. 677 – 690, May 2004.

[233] J. E. Loughney, "Context Transfer Protocol," Internet Draft, Dec. 2005, draft-ietf-seamoby-ctp-11.txt.

[234] M. Liebsch *et al.*, "Candidate Access Router Discovery," Internet Draft, Sep. 2004, draft-ietf-seamoby-card-protocol-08.txt.

[235] J. Kempf and Ed., "Problem Description: Reasons For Performing Context Transfers Between Nodes in an IP Access Network," Internet Engineering Task Force, RFC 3374, Sept. 2002.

[236] A. Rojas and J. Paradells, "Interference and Capacity in CDMA Hybrid Systems with Hysteresis Margin," in *Proceedings of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 1998.

[237] Y.-P. Wang and T. Ottosson, "Cell Search in W-CDMA," *IEEE Journal in Selected Areas Communications*, vol. 18, no. 8, pp. 1470–1482, Aug. 2000.

[238] A. Acampora and M. Naghshineh, "An Architecture and Methodology for Mobile-executed handoff in Cellular ATM Networks," in *IEEE Journal on Selected Areas of Communication*, vol. vol. 12(U), Oct. 1994, pp. 1365–1375.

[239] T. Pagtzis, "Interfacing Mediation Services between KQML and APRIL through SMTP," Master's thesis, Imperial College, University of London, Sep. 1996.

[240] A. Hatami and K. Pahlavan, "A Comparative Performance Evaluation of RSS-based Positioning Algorithms used in WLAN Networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2003.

[241] A. Ghosh, D. Wolter, J. Andrews, and R. Chen, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 129–136, 2004.

[242] A. Fox, S. Gribble, Y. Chawathe, E. Brewer, and P. Gauthier, "Cluster-based scalable network services," in *Proceedings of Symposium on Operating Systems Principles*, Jun. 1999.

[243] K. Carlberg and R. Atkinson, "General Requirements for Emergency Telecommunication Service (ETS)," Internet Engineering Task Force, RFC 3689, Feb. 2004.

[244] K. Kuppuswamy and D. Lee, "On Subscription Admission Control for Network Service Provisioning," *IEEE Communications Letters*, vol. 9, no. 1, pp. 66–68, Jan. 2005.

[245] D. Funato *et al.*, "Geographically Adjacent Access Router Discovery Protocol," Internet Draft, Apr. 2002, draft-funato-mipshop-gaardp-00.txt.

[246] C. Westphal and R. Koodli, "IP Header Compression: a Study of Context Establishment," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2003.

[247] C. Bormann, C. Burmeister, M. Degermark, H. Fukushima, H. Hannu, L.-E. Jonsson, R. Hakenberg, T. Koren, K. Le, Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, and H. Zheng, "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed," Internet Engineering Task Force, RFC 3095, July 2001.

[248] J. Tourrilhes, "Wireless Tools for Linux," Apr. 1996. [Online]. Available: http://www.hpl.hp.com/personal/Jean_Tourrilhes/Linux/Tools.html

[249] J. Banks, J. Carson, and B. Nelson, *Discrete Event System Simulation.* Prentice Hall, 2002.

[250] ISI, "The Network Simulator - NS-2," http://www.isi.edu/nsnam/ns/, 1989.

[251] E. Zegura, K. Calvert, and S. Bhattacharjee, "How to Model an Internetwork," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Mar. 1996.

[252] S. Floyd and V. Paxson, "Difficulties in Simulating the Internet," *IEEE/ACM Transactions on Networking*, vol. 9, no. 4, pp. 392–403, Aug. 2001.

[253] K. Calvert, M. Doar, and E. Zegura, "Modeling internet topology," *IEEE Communications Magazine*, vol. 35, no. 6, pp. 160–163, Jun. 1997.

[254] S. S. H. Tangmunarunkit, R. Govindan, "Internet Path Inflation Due to Policy Routing," in *Proceedings of SPIE ITCom*, Aug. 2001.

[255] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network Topology Generators: Degree-based vs. Structural," *ACM Computer Communication Review*, vol. 32, no. 4, pp. 147–159, Aug. 2002.

[256] A. Broido and K. Claffy, "Internet topology: connectivity of IP graphs," in *Proceedings of SPIE conference*, Aug. 2001.

[257] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," *ACM Computer Communication Review*, vol. 29, no. 4, pp. 251–262, Aug. 1999.

[258] R. Albert and A. Barabasi, "Topology of Evolving Networks: Local events and Universality," *Physical Review Letters*, vol. 85, no. 2, pp. 5234–5237, May 2000.

[259] A. Medina, I. Matta, and J. Byers, "On the Origin of Power Laws in Internet Topologies," *ACM Computer Communication Review*, vol. 30, no. 2, Apr. 2000.

[260] B. Waxman, "Routing of Multi-point Connections," *IEEE Journal in Selected Areas Communications*, vol. 6, no. 9, pp. 1617–1622, Dec. 1988.

[261] G. Tech, "GT-ITM. Georgia Tech Internetwork Topology Models," www.cc.gatech.edu/projects/gtitm/, 1997.

[262] ISI, "TIERS. Tiers Topology Generator," http://www.isi.edu/nsnam/ns/ns-topogen.html#tiers, 1995.

[263] J. Winick and S. Jamin, *Inet-3.0: Internet topology generator*, University of Michigan, 2002.

[264] University of Michigan, "Inet Topology Generator," http://topology.eecs.umich.edu/inet/, 1996.

[265] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRITE: an Approach to Universal Topology Generation," in *Proceedings of 9th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, Aug. 2001.

[266] A. Medina, I. Matta, and J. Byers, "BRITE. Boston University Representative Internet Topology Generator," 2001.

[267] J. Chen, D. Gupta, K. Vishwanath, A. Snoeren, and A. Vahdat, "Routing in an Internet-scale Network Emulator," in *Proceedings of Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS)*, Oct. 2003.

[268] J. Guillaume and M. Latapy, "A Realistic Model for Complex Networks," http://arxiv.org/pdf/cond-mat/0307095.

[269] E. Dijkstra, "A note on two problems in conneciton with graphs," *Numerical Mathematics*, vol. 2, no. 1, pp. 269–271, 1959.

[270] J. Morrow, "Polarization-adjustable Omni-directional Dipole Array," *Antennas and Wireless Propagation Letters*, vol. 2, no. 1, pp. 223–225, 2003.

[271] J. Weitzen and M. Wallace, "Analysis of Diversity Performance of Space Diversity and Cross Polarization for PCS Base Stations," in *Proceedings of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 1998.

[272] K.-L. W. F.-R. H. C.-L. Tang, "A low-profile Omni-directional Circularly Polarized Antenna for WLAN Access Point," in *Proceedings of Antennas and Propagation Society Symposium*, Jun. 2004.

[273] S. Nanda, "Teletraffic Models for Urban and Suburban Microcells: Cell Sizes and Handoff Rates," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 673 – 682, Nov. 1993.

[274] D. Bansal, A. Chandra, R. Shorey, A. Kulshreshtha, and M. Gupta, "Mobility Models for Cellular systems: Cell Topography and Handoff Probability," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Spring)*, May 1999.

[275] J. Yoon, M. Liu, and B. Noble, "Sound Mobility Models," in *Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, Apr. 2003.

[276] L. B. J.Y. and M. Vojnovi, "Perfect Simulation and Stationarity of a Class of Mobility Models," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Jul. 2005.

[277] J. Yoon, M. Liu, and B. Noble, "Random Way-point considered Harmful," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2003.

[278] S. Brown, C. Bryett, and M. Mowle, "In-Service Emissions Performance-Drive Cycles Vol.I," http://www.ephc.gov.au/pdf/diesel/diesel_Project2_1.pdf, National Environment Protection Council (NEPC), Tech. Rep., Mar. 1999.

[279] ——, "In-Service Emissions Performance-Drive Cycles Vol.II," http://www.ephc.gov.au/pdf/diesel/Vol2_Attachment1.pdf, National Environment Protection Council (NEPC), Tech. Rep., May 1999.

[280] T. Pagtzis, R. Chakravorty, J. Crowcroft, S. Hailes, and P. Kirstein, "Proactive Mobile IPv6 for Context-aware all-IP Wireless Access Networks," in *Proceedings of Annual IEEE Conference on Wireless Communications*, Jun. 2005.

[281] T. Aura, "Cryptographically Generated Addresses (CGA)," Internet Engineering Task Force, RFC 3972, Mar. 2005.

[282] R. Chalmers, "Improving Device Mobility with Intelligence at the Network Edge," Ph.D. dissertation, University of California, Santa Barbara, Sep. 2003.

[283] R. Vijayan and J. M. Holtzman, "A Model for analyzing Handoff Algorithms,," *Transactions on Vehicular Technology*, p. 351356, Aug. 1993.

[284] J. Chen and H. Lin, "A Gateway Approach to Mobility Integration of GPRS and Wireless LANs," *IEEE Wireless Communications*, vol. 12, no. 2, pp. 86–95, Apr. 2005.

[285] M. Akar and U. Mitra, "Optimal Handoff Control: incorporating Handoff Delay," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Sep. 2000, pp. 484 - 489.

[286] S. Das, W. MacDonald, and H. Viswanathan, "Sensitivity Analysis of Handoff Algorithms on CDMA Forward Link," *IEEE Journal in Selected Areas Communications*, pp. 272–285, Jan. 2005.

[287] H. Balakrishnan, S. Seshan, and R. Katz, "Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks," *Wireless Networks*, vol. 1, no. 4, Feb. 1995.

[288] A. Bakre and B. Badrinath, "I-TCP: Indirect TCP for Mobile Hosts," in *15th IEEE Conference on Distributed Computing Systems*, May 1995.

[289] K. Brown and S. Singh, "M-tcp: Tcp for mobile cellular networks." *ACM Computer Communication Review*, vol. 27, no. 2, 1997.

[290] V. Bakshi, P. Krishna, N. Vaidya, and D. Pradhan, "Improving performance of tcp over wireless networks," in *Proceedings of 17th IEEE Conference on Distributed Computing Systems*, May 1997.

[291] J. Hu and K. L. Yeung, "A novel base station flow control scheme for tcp over heterogeneous networks," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Jun. 2001.

[292] K. Ratnam and I. Matta, "WTCP: An Efficient Mechanism for Improving TCP Performance over Wireless Links," in *Proceedings of IEEE Symposium on Computer and Communications*, Mar. 1998.

[293] R. Chakravorty, S. Katti, J. Crowcroft, and I. Pratt, "Flow aggregation for enhanced tcp over wide-area wireless," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2003.

[294] C. Sreenan, J.-C. Chen, P. Agrawal, and B. Narendran, "Delay Reduction Techniques for Playout Buffering," *IEEE Transactions on Multimedia*, vol. 2, no. 2, pp. 88–100, Jun. 2000.

[295] M. Kalman, E. Steinbach, and B. Girod, "Adaptive Media Playout for Low-Delay Video Streaming over Error-prone Channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 841 – 851, Jun. 2004.

[296] N. Kitawaki and K. Itoh, "Pure Delay Effects on Speech Quality in Telecommunications," *IEEE Journal in Selected Areas Communications*, vol. 9, no. 4, pp. 586 – 593, 1991.

[297] H. Sanneck, A. Stenger, K. Ben Younes, and B. Girod, "A New Technique for Audio Packet Loss Concealment," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, Sep. 1996, pp. 48 – 52.

[298] Y. Liang, N. Farber, and B. Girod, "Adaptive Playout Scheduling and Loss Concealment for Voice Communication over IP Networks," *IEEE Transactions on Multimedia*, vol. 5, no. 4, pp. 532 – 543, Dec. 2003.

[299] K. Malki and H. Soliman, "Simultaneous Bindings for Mobile IPv6 Fast Handoffs," Internet Draft, Jun. 2002, draft-elmalki-mobileip-bicasting-v6-02.txt.

[300] K. Malki, "Low latency Handoffs in Mobile IPv4," Internet Draft, June 2004, draft-ietf-mobileip-lowlatency-handoffs-v4-09.txt.

[301] H. Soliman, C. Castelluccia, K. Malki, and L. Bellier, "Hierarchical Mobile IPv6 mobility management (HMIPv6)," Internet Draft, Dec. 2004, draft-ietf-mipshop-hmipv6-04.txt.

[302] J. Kim, G. Stuber, I. Akyildiz, and B. Chung, "Soft Handoff Analysis of Hierarchical CDMA Cellular Systems," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 3, pp. 1122–1134, May 2005.

[303] A. Mishra, M. Shin, and W. Arbaugh, "An empirical analysis of the ieee 802.11 mac layer handoff process," *ACM Computer Communication Review*, vol. 33, no. 2, pp. 93–102, 2003.

[304] I. Rasmussen, L.K.and Oppermann, "Ping-Pong Effects in Linear Parallel Interference Cancellation for CDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 357–363, Mar. 2003.

[305] G. Fan, I. Stojmenovic, and J. Zhang, "A Triple-layer Location Management Strategy for Wireless Cellular Networks," in *Proceedings of IEEE International Conference on Communications (ICC)*, Oct. 2002.

[306] T. Henderson, D. Kotz, and I. Abyzov, "Experimental Testbeds and Data: The Changing Usage of a Mature Campus-wide Wireless Network," in *Proceedings*

of *ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, Sep. 2004.

[307] P. Zerfos, G. Zhong, J. Cheng, H. Luo, S. Lu, and J. Jia-Ru Li, "DIRAC: a software-based wireless Router System," in *Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, Sep. 2003.

[308] R. Hinden and S. Deering, "Internet Protocol Version 6 (IPv6) Addressing Architecture," Internet Engineering Task Force, RFC 3513, Apr. 2003.

[309] R. Chakravorty, I. Pratt, and J. Crowcroft, "A Framework for Dynamic SLA-based QoS Control for UMTS," *IEEE Wireless Communications*, vol. 10, no. 5, pp. 30–37, Oct. 2003.

[310] R. Chakravorty, I. Pratt, J. Crowcroft, and M. D'Arienzo, "Dynamic SLA-based QoS control for Third Generation Wireless Networks: the CADENUS extension," in *Proceedings of IEEE International Conference on Communications (ICC)*, May 2003.

[311] B. Haberman and D. Thaler, "Unicast-Prefix-based IPv6 Multicast Addresses," Internet Engineering Task Force, RFC 3306, Aug. 2002.

[312] D. Ralph, J. Durand, P. O'Hanlon, *et al.*, "IPv6 Multicast Address Allocation Study," Renater, Tech. Rep. Apr., 2003.

[313] R. Hinden, M. O'Dell, and S. Deering, "An IPv6 Aggregatable Global Unicast Address Format," Internet Engineering Task Force, RFC 2374, July 1998.

[314] IEEE, *Guidelines for 64-bit Global Identifier (EUI-64) Registration Authority*, http://standards.ieee.org/db/oui/tutorials/EUI64.html, Mar. 1997.

[315] D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, and L. Wei, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification," Internet Engineering Task Force, RFC 2362, June 1998.

[316] A. Ballardie, "Core Based Trees (CBT) Multicast Routing Architecture," Internet Engineering Task Force, RFC 2201, Sept. 1997.

[317] L. Wei and D. Estrin, "The Trade-offs of Multicast Trees and Algorithms," in *Proceedings of International Conference on Computer Communications and Networks*, Sep. 1994.

[318] D. Estrin, M. Handley, A. Helmy, P. Huang, and D. Thaler, "A dynamic bootstrap mechanism for rendezvous-based multicast routing," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Mar. 1999, pp. 1090 – 1098.

[319] D. Thaler and C. Ravishankar, "Distributed Center-Location Algorithms," *IEEE Journal in Selected Areas Communications*, vol. 15, no. 3, pp. 291–303, Apr. 1997.

[320] H. Hsieh and R. Sivakumar, "On using Peer-to-Peer Communication in Cellular Wireless Data Networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 1, pp. 57–72, Jan. 2004.

[321] S. Deering, W. Fenner, and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6," Internet Engineering Task Force, RFC 2710, Oct. 1999.

[322] B. Cain, S. Deering, I. Kouvelas, B. Fenner, and A. Thyagarajan, "Internet Group Management Protocol, Version 3," Internet Engineering Task Force, RFC 3376, Oct. 2002.

[323] W. Fenner, "Internet Group Management Protocol, Version 2," Internet Engineering Task Force, RFC 2236, Nov. 1997.

[324] S. Deering, "Host extensions for IP multicasting," Internet Engineering Task Force, RFC 1112, Aug. 1989.

[325] R. Lupas and S. Verd, "Nearfar resistance of multiuser detectors in asynchronous channels," *IEEE Transactions on Communications*, vol. 38, no. 1, p. 496508, Apr. 1990.

[326] S. Moshavi, "Multi-user Detection for DS-CDMA Communications," *IEEE Communications Magazine,*, vol. 34, no. 10, pp. 124 – 136, 1996.

[327] D. Guo, L. Rasmussen, and T. Lim, "Linear Parallel Interference Cancellation in Long-code CDMA Multiuser Detection," *IEEE Journal in Selected Areas Communications*, vol. 17, no. 3, p. 20742081, Dec. 1999.

[328] T. S. Rappaport, *Wireless Communications, Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[329] H. Maturino-Lozoya, D. Munoz-Rodriguez, F. Jaimes-Romera, and H. Tawfik, "Handoff Algorithms based on Fuzzy Classifiers," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 6, pp. 2286 – 2294, Nov. 2000.

[330] A. Murase, I. Symington, and E. Green, "Handover Criterion for Macro- and Micro-cellular Systems," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Spring)*, May 1991.

[331] A. Rojas and J. Paradells, "Shadow Margin, Overlapping and Transceivers per cell when making Handovers with Hysteresis Margin," in *Proceedings of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep. 1998.

[332] T. R. WGUWG4, "Technical Specifications: Criteria for Soft Handover Algorithm," 3GPP, Tech. Rep. TSG R2#4(99)378, May 1999.

[333] B. Schneier, *Applied Cryptography*, 2nd ed. John Wiley & Sons, 1996.

[334] B. Cherkassky, A. Goldberg, and T. Radzik, "Shortest Paths Algorithms: Theory and Experimental Evaluation," *Mathematical Programming*, vol. 73, p. 129174, 1996.

[335] I. Ramani and S. Savage, "Syncscan: Practical fast handoff for 802.11 infrastructure networks," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Miami, FL, Mar. 2005.

[336] R. Chandra and P. Bahl, "MultiNet: Connecting to Multiple IEEE 802.11 Networks using a Single Wireless Card," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Mar. 2004.

[337] R. Koodli and C. Perkins, "Fast Handovers and Context Transfers in Mobile Networks," *ACM Computer Communication Review*, vol. 31, no. 5, pp. 37–47, Oct. 2001.

[338] J. Kempf, P. Calhoun, G. Dommety, S. Thalanany, A. Singh, P. McCann, and T. Hiller, "Bidirectional Edge Tunnel Handover for IPv6," Internet Draft, Dec. 2001, draft-kempf-beth-ipv6-00.txt.

[339] G. Tsirtsis, (Ed.), A. Yegin, C. Perkins, G. Dommety, K. Malki, and M. Khalil, "Fast Handovers for Mobile IPv6," Internet Draft, Apr. 2001, draft-ietf-mobileip-fast-mipv6-01.txt.

[340] G. Dommety, (Ed.), A. Yegin, C. Perkins, G. Tsirtsis, K. Malki, and M. Khalil, "Fast Handovers for Mobile IPv6," Internet Draft, Mar. 2002, draft-ietf-mobileip-fast-mipv6-04.txt.

[341] T. Pagtzis, "A Model for Proactive Seamless Mobility in IPv6 networks," University College London, London, UK, Technical Report UCL/CS-TR-01-133, Jan. 2001.

[342] G. Dommety, (Ed.), A. Yegin, C. Perkins, G. Tsirtsis, K. Malki, and M. Khalil, "Fast Handovers for Mobile IPv6," Internet Draft, Jul. 2001, draft-ietf-mobileip-fast-mipv6-02.txt.

[343] ——, "Fast Handovers for Mobile IPv6," Internet Draft, Jul. 2001, draft-ietf-mobileip-fast-mipv6-03.txt.

[344] R. Koodli, (Ed.), *et al.*, "Fast Handovers for Mobile IPv6," Internet Draft, Sep. 2002, draft-ietf-mobileip-fast-mipv6-05.txt.

[345] ——, "Fast Handovers for Mobile IPv6," Internet Draft, Mar. 2003, draft-ietf-mobileip-fast-mipv6-06.txt.

[346] ——, "Fast Handovers for Mobile IPv6," Internet Draft, Sep. 2003, draft-ietf-mobileip-fast-mipv6-07.txt.

[347] ——, "Fast Handovers for Mobile IPv6," Internet Draft, Oct. 2003, draft-ietf-mobileip-fast-mipv6-08.txt.

[348] ——, "Fast Handovers for Mobile IPv6," Internet Draft, Oct. 2003, draft-ietf-mipshop-fast-mipv6-00.txt.

[349] ——, "Fast Handovers for Mobile IPv6," Internet Draft, Jan. 2004, draft-ietf-mipshop-fast-mipv6-01.txt.

[350] ——, "Fast Handovers for Mobile IPv6," Internet Draft, Jul. 2004, draft-ietf-mipshop-fast-mipv6-02.txt.

[351] G. Krishanmurti, "Requirements for CAR Discovery Protocols," Internet Draft, Aug. 2002, draft-krishanmurti-mipshop-car-req-00.txt.

[352] D. Trossen, G. Krishnamurthi, H. Chaskar, R. Chalmers, and E. Shim, "A Dynamic Protocol for Candidate Access-Router Discovery," Internet Draft, Sep. 2004, draft-trossen-seamoby-dycard-00.txt.

[353] M. Liebsch, Ed., A. Singh, Ed., H. Chaskar, D. Funato, and E. Shim, "Candidate Access Router Discovery (CARD)," Internet Engineering Task Force, RFC 4066, July 2005.

[354] P. McCann, "Mobile IPv6 Fast Handovers for 802.11 Networks," Internet Draft, Oct. 2004, draft-ietf-mipshop-80211fh-03.txt.

[355] H. Edelsbrunner, "The Union of Balls and its Dual Shape," in *Proceedings of the 9th Annual Symposium on Computational geometry*, Jul. 1993.

[356] A. Ransom, "Handoff Considerations in Microcellular Systems Planning," in *Proceedings of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Aug. 1995, pp. 804 –808.

[357] J. Wigard, T. Nielsen, P. Michaelsen, and P. Morgensen, "On a Handover Algorithm in a PCS1900/GSM/DCS1800 network," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Spring)*, Apr. 1999, pp. 2510 – 2514.

[358] G. Liu and G. J. Maguire, "A Class of Mobile Motion Prediction Algorithms for Wireless Mobile Computing and Communication," *ACM Journal in Mobile Networking and Applications*, pp. 113 – 121, Oct. 1996.

[359] F. Feng and D. Reeves, "Explicit Proactive Handoff with Motion Prediction for Mobile IP," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2004.

[360] Y. Fei and V. Leung, "Mobility-based Predictive Call-admission Control and Bandwidth Reservation in Wireless Cellular Networks," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2001.

[361] W.-S. Soh and H. Kim, "QoS provisioning in Cellular Networks based on Mobility Prediction Techniques," *IEEE Communications Magazine*, vol. 41, no. 1, pp. 86 – 92, Jan. 2003.

[362] S. Thompson, "Intelligent Hub Access System," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, Oct. 2000.

[363] W. Moh, L. Xiang, X. Zhao, and E. Park, "Differentiated-service-based Inter-domain Multicast Routing: Enhancement of MBGP," in *Proceedings of 9th International Conference on Computer Communications and Networks*, Oct. 2000.

[364] T. Bates, R. Chandra, D. Katz, and Y. Rekhter, "Multiprotocol Extensions for BGP-4," Internet Engineering Task Force, RFC 2283, Feb. 1998.

[365] S. Dimolitsas and J. Phipps, "Experimental Quantification of Voice Transmission Quality of Mobile-Satellite Personal Communications systems," *IEEE Journal in Selected Areas Communications*, vol. 13, no. 2, pp. 458 – 464, 1995.

[366] . WG, *IEEE 802.11a Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification Standard: High Speed Physical Layer in the 5 GHz band*, IEEE, Oct. 1999.

[367] A. Gyasi-Agyei, "Mobile IP-DECT Internetworking Architecture supporting IMT-2000 applications," *IEEE Network*, vol. 15, no. 6, pp. 10 – 22, 2001.

[368] D. Johnson, "Scalable Support for Transparent Mobile Host Internetworking," *Wireless Networks*, vol. 3, no. 1, pp. 311–321, Mar. 1995.

[369] D. B. Johnson and C. Perkins, "Route Optimization in Mobile IP," Internet Draft, Mar. 1996, draft-ietf-mobileip-optim-10.txt.

[370] T. Ernst, "Network Mobility Support Goals and Requirements," Internet Draft, Feb. 2005, draft-ietf-nemo-requirements-04.txt.

[371] W. Winoto, E. Schwartz, H. Balakrishnan, and J. Lilley, "The Design and Implementation of an Intentional Naming System," in *Proceedings of the seventeenth ACM symposium on Operating systems principles*, Dec. 1999.

[372] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed Diffusion for Wireless Sensor Networking," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 2 – 16, 2003.

[373] C. J. Navas and T. Imielinski, "Geocast-geographic addressing and routing," in *MOBICOM '97.* ACM, Sept. 1997.

[374] C. E. Williams, "Localized Mobility Management Goals," draft-ietf-mipshop-lmm-requirements-03, Internet Engineering Task Force, Internet Draft (Work in progress), Jan. 2004.

[375] G. Kirby, "Locating the User," in *Communications International*, Oct. 1995.

[376] R. Caceres and V. Padmanabhan, "Fast and Scalable Wireless Handoffs in Support of Mobile Internet Audio," *ACM Journal in Mobile Networking and Applications*, vol. 3, no. 4, pp. 351–363, Jun. 1998.

[377] E. Gustafsson, "Mobile IPv4 Regional Registration," Internet Draft, Nov. 2004, draft-ietf-mip4-reg-tunnel-00.txt.

[378] C. Castelluccia, "A Hierarchical Mobility Management Scheme for the Internet," Stanford University, Technical Report CSL-TR-97-736, Sep. 1997.

[379] Z. Hsieh, R.and Zhou and A. Seneviratne, "S-MIP: a Seamless Handoff Architecture for Mobile IP," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2003.

[380] A. Campbell, J. Gomez, S. Kim, *et al.*, "Comparison of IP Micro-mobility protocols," *IEEE Wireless Communications*, vol. 9, no. 1, pp. 72–82, 2002.

[381] K. Varadhan, L. Salgarelli, T. La Porta, R. Ramjee, and S. Thuel, "IP micro-mobility support using HAWAII," Internet Draft, Jul. 2000, draft-ietf-mobileip-hawaii-01.txt.

[382] P. Reinbold and O. Bonaventure, "A Comparison of Mobility Protocols"," University of Namur," Technical Report Infonet-2001-07, Jun. 2001.

[383] K. D. Wong, H. Wei, A. Dutta, K. Young, and H. Schulzrinne, "Performance of IP Micro-Mobility Management Scehemes using Host-based Routing," in *Proceedings of Wireless Personal Multimedia Communications*, Sep. 2001.

[384] N. Maxemchuk and S. Low, "Active Routing," *IEEE Journal in Selected Areas Communications*, vol. 19, no. 3, pp. 552–565, Mar. 2001.

[385] M. Eltahir and J. Dunlop, "Cellular Layout Issues on Performance of Cellular IP," *Electronics Letters*, vol. 39, no. 12, pp. 942 – 944, Jun. 2003.

[386] J. Castellanos, S. Kim, Z. Turanyi, *et al.*, "A Comparison of IP-Micromobility Protocols," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 4, no. 4, pp. 45–53, Oct. 2000.

[387] A. Grilo, P. Estrela, and M. Nunes, "Terminal independent mobility for ip (timip)," *IEEE Computer Communications*, vol. 39, no. 12, pp. 34–41, Dec. 2001.

[388] A. O'Neil, S. Corson, and G. Tsirtsis, "Routing and Handoff in Edge Mobility Architecture," Flarion Technologies, Tech. Rep., Jun. 2000.

[389] A. Mihailovic and P. Eardley, "A Framework for Evaluation of IP Mobility Protocols," in *Proceedings of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2000, pp. 451–457.

[390] A. O'Neill, S. Corson, and G. Tsirtsis, "Edge Mobility Architecture," Internet Draft (Work in progress), Jul. 2000, draft-oneill-ema-02.txt.

[391] S. Corson and V. Park, "Temporally-Ordered Routing Algorithm (TORA) Version 1 Functional Specification," Internet Draft (Work in progress), Nov. 2000, draft-ietf-manet-tora-spec-03.txt.

[392] R. Callon, "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments," RFC, Internet Engineering Task Force, RFC 1195, Dec. 1990.

[393] J. Moy, "OSPF Version 2," Internet Engineering Task Force, RFC 2178, July 1997.

[394] S. Das, A. Misra, and P. Agrawal, "TeleMIP: Telecommunications-enhanced Mobile-IP Architecture for Fast Intra-domain Mobility," *IEEE Personal Communications*, vol. 7, no. 4, pp. 50–58, Aug. 1998.

[395] A. Helmy, "A Multicast-based Protocol for IP Mobility Support," in *ACM 2nd International Workshop on Networked Group Communication (NGC)*, Nov. 2000.

[396] D. Saha, A. Mukherjee, I. Misra, M. Chakraborty, and N. Subhash, "Mobility support in IP: a Survey of related Protocols," *IEEE Network*, vol. 18, no. 6, pp. 34–40, Dec. 2004.

[397] A. Misra, S. Das, A. Mcauley, A. Dutta, and S. Das, "Integrating QoS support in TeleMIP's Mobility Architecture," in *Proceedings of IEEE International Conference on Personal Wireless Communications*, Dec. 2000, pp. 57 – 64.

[398] A. Misra, S. Das, A. McAuley, and S. Das, "Autoconfiguration, Registration, and Mobility Management for Pervasive Computing," *IEEE Personal Communications*, vol. 8, no. 4, pp. 24–31, 2001.

[399] M. Dunmore and T. Pagtzis, "Mobile ipv6 handovers: Performance analysis and evaluation," Lancaster/UCL-6NET, Tech. Rep. 32603/ULANC/DS/4.1.1/A1, May 2004.

[400] Y. Qiu, J. Zhou, and F. Bao, "Protecting all Traffic Channels in mobile IPv6 Network," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2004.

[401] R. Droms, Ed., J. Bound, B. Volz, T. Lemon, C. Perkins, and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)," Internet Engineering Task Force, RFC 3315, July 2003.

[402] B. P. G. Daley and R. Nelson, "Effects of Fast Router Advertisement on Mobile IPv6 Handovers," in *Proceedings of The Eighth IEEE Symposium on Computers and Communications (ISCC)*, Jun. 2003.

[403] G. Daley, B. Pentland, and R. Nelson, "Movement Detection Optimizations in Mobile IPv6," in *Proceedings of The 11th IEEE International Conference on Networks (ICON)*, Sep. 2003.

[404] R. Chakravorty, P. Vidales, K. Subramanian, I. Pratt, and J. Crowcroft, "Practical Experience with Wireless Networks Integration using Mobile IPv6," *ACM Computer Communication Review*, vol. 3, no. 4, Oct. 2003.

[405] Y. Park and S. Rappaport, "Performance Analysis of Session-oriented Data Communications for Mobile Computing in Cellular Systems," *Wireless Networks*, vol. 6, no. 6, pp. 441–456, Dec. 2000.

[406] F. Borgonovo, A. Capone, L. Fratta, and M. Marchese, "MAC for WATM Air Interface: Impact of Error Control Schemes on Protocol Design," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Spring)*, May 1999.

[407] S. Choi and K. Shin, "A Class of Adaptive Hybrid ARQ Schemes for Wireless Links," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 3, pp. 777–790, May 2001.

[408] A. Adya, P. Bahl, R. Chandra, and L. Qiu, "Architecture and Techniques for Diagnosing Faults in IEEE 802.11 Infrastructure Networks," in *Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, Sep. 2004.

[409] A. Mishra, M. Shin, and W. Arbaugh, "Context Caching using Neighbor-graphs for Fast Handoffs in a Wireless Network," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Mar. 2004.

[410] S. Shin, A. Forte, A. Rawat, and H. Schulzrinne, "Reducing MAC Layer Handoff Latency in IEEE 802.11 wireless LANs," in *Proceedings of the 2nd International workshop on Mobility management & Wireless Access Protocols*, Oct. 2004.

[411] A. Ganz, A. Phonphoem, N. Llopis, I. Kim, K. Wongthavarawat, and Z. Ganz, "Converged Voice, Video and Data Wired/Wireless LANs Testbed," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, 1999, pp. 1297–1301.

[412] G. Anastasi and L. Lenzini, "QoS provided by the IEEE 802.11 Wireless LAN to Advanced Data Applications: a Simulation Analysis," *ACM/Baltzer Wireless Networks Journal*, 2000.

[413] T. Suzuki and S. Tasaka, "Performance Evaluation of Integrated Video and Data Transmission with the IEEE 802.11 Standard MAC Protocol," in *Proceedings of IEEE Global Communications Conference (GLOBECOM)*, vol. 1b, Dec. 1999, pp. 580 –586.

[414] A. Lindgren, A. Almquist, and O. Schelen, "Evaluation of quality of service schemes for IEEE 802.11 wireless LANs," in *Proceedings of IEEE Conference on Local Computer Networks (LCN)*, Nov. 2001, pp. 348–351.

[415] L. Wells, S. Christensen, L. Kristensen, and K. Mortensen, "Simulation-based Performance Analysis of Web Servers," in *Proceedings of 9th International Workshop on Petri Nets and Performance Models*, Sep. 2001, pp. 59–68.

[416] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview," Internet Engineering Task Force, RFC 1633, June 1994.

[417] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Service," Internet Engineering Task Force, RFC 2475, Dec. 1998.

[418] C. Simmonds and M. Beach, "Network Planning aspects of DS-CDMA with particular emphasis on Soft Handoff," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Spring)*, May 1993.

[419] E. Bergeron, C. Despins, and G. Delisle, "Impact of Radio-link Parameters on the Capacity of UHF and Millimeter-wave Indoor Cellular Systems," *IEEE Transactions on Vehicular Technology*, vol. 46, no. 1, pp. 21–30, Feb. 1997.

[420] T. S. Kim, H.-S. Cho, and D. K. Sung, "Moving Elevator-cell System in Indoor Buildings," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1743–1751, Sep. 2000.

[421] M. Ergen, S. Coleri, B. Dundar, R. Jain, A. Puri, and P. Varaiya, "Application of GPS to Mobile IP and Routing in Wireless Networks," in *Proceedings of IEEE The Semi-annual Vehicular Technology Conference (VTC-Fall)*, Sep. 2002.

[422] F. Harary, *Graph Theory.* Addison-Wesley, 1969.

[423] G. Di Caro and M. Dorigo, "Distributed Stigmergetic Control for Communications Networks," *Joumal of Artificial Intelligence Research*, vol. 9, no. 1, pp. 317–365, May 1998.

[424] B. Aboba and M. Beadles, "The Network Access Identifier," Internet Engineering Task Force, RFC 2486, Jan. 1999.

[425] M. Rahnema, "Overview of the GSM system and Protocol Architecture," *IEEE Communications Magazine*, vol. 31, no. 4, pp. 92–100, Apr. 1993.

[426] ETSI, *GSM 02.17: Digital cellular telecommunications system (Phase 2+); Subscriber Identity Modules (SIM); Functional characteristics.*

[427] R. K. Jain, *The Art of Computer Systems Performance Analysis : Techniques for Experimental Design, Measurement, Simulation, and Modeling.* Wiley, 1991.

[428] P. Dalgaard, *Introductory Statistics with R.* Springer, Aug. 2002.

[429] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive Playout Mechanisms for Packetized Audio Applications in Wide-area Networks," in *Proceedings of Networking for Global Communications*, Jun. 1994.

[430] S. Moon, J. Kurose, and D. Towsley, "Packet Audio Playout Delay Adjustment: Performance Bounds and algorithms," *ACM/Springer Multimedia Systems*, pp. 17–28, 1998.

[431] J. Pinto and K. Christensen, "An Algorithm for Playout of Packet Voice based on Adaptive Adjustment of Talkspurt Silence Periods," in *Proceedings of Conference on Local Computer Networks*, Oct. 1999.

[432] J. L. Q. Rosenberg and H. Schulzrinne, "Integrating Packet FEC into Adaptive Voice Playout Buffer Algorithms on the Internet," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, Mar. 2000.

[433] J. Holub, J. Beerends, and R. Smid, "A Dependence between Average Call Duration and Voice Transmission Quality: Measurement and Applications," in *Proceedings of Wireless Telecommunications Symposium*, May 2004, pp. 75–81.

[434] B. Gavish and S. Sridhar, "The Impact of Mobility on Cellular Network Configuration," *Wireless Networks*, vol. 7, no. 2, pp. 173–185, Apr. 2001.

[435] N. Feuersteihn, M.; Sollenberger, "Experimental Tests of Coverage and Handoff for a Low-power TDMA Wireless Access System," in *3rd Annual International Conference on Universal Personal Communications*, Sep. 1994, pp. 501 - 505.

[436] J. Sarnecki, C. Vinodrai, A. Javed, P. O'Kelly, and K. Dick, "Microcell Design Principles," *IEEE Communications Magazine*, vol. 31, no. 4, pp. 76–82, Apr. 1993.

[437] H. Velayos and G. Karlsson, "Techniques to Reduce the IEEE 802.11b Handoff Time," in *Proceedings of IEEE International Conference on Communications (ICC)*, Jun. 2004.

[438] S. Sharma, N. Zhu, and T. Chiueh, "Low-latency Mobile IP Handoff for Infrastructure-mode Wireless LANs," *IEEE Journal in Selected Areas Communications*, vol. 22, no. 4, pp. 643–652, 2004.

[439] M. Carvalho and J. Garcia-Luna-Aceves, "Delay Analysis of IEEE 802.11 in Single-hop Networks," in *Proceedings of 11th IEEE International Conference on Network Protocols*, Nov. 2003, pp. 146–155.

[440] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal in Selected Areas Communications*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[441] M. Panda, A. Kumar, and S. Srinivasan, "Saturation Throughput Analysis of a System of Interfering IEEE 802.11 WLANs," in *Proceedings of World of Wireless Mobile and Multimedia Networks*, Jun. 2005.

[442] X. Yang, "Backoff-based priority schemes for IEEE 802.11," in *Proceedings of IEEE International Conference on Communications (ICC)*, May 2003.

[443] M. Manshaei, G. Cantieni, C. Barakat, and T. Turletti, "Performance Analysis of the IEEE 802.11 MAC and Physical Layer Protocol," in *Proceedings of World of Wireless Mobile and Multimedia Networks*, Jun. 2005.

[444] Z. Hadzi-Velkov and B. Spasenovski, "Saturation Throughput - Delay Analysis of IEEE 802.11 DCF in Fading Channel," in *Proceedings of IEEE International Conference on Communications (ICC)*, May 2003.

[445] D. Bantz and F. Bauchot, "Wireless LAN Design Alternatives," *IEEE Network*, vol. 8, no. 2, Mar. 1994.

[446] V. Srinivasan, K. Sivarajan, and H. Jiang, "Channel Allocation in Tiered Cellular Networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Sep. 1999.

[447] W. H. Beyer, Ed., *CRC Standard Mathematical Tables*, 28th ed. Boca Raton, FL: CRC Press, 1987.